European University Institute

DEPARTMENT OF ECONOMICS

EUI WORKING PAPERS

ECO 2016/09
Department of Economics

Education, HIV Status and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter ?

Daniela Iorio and Raül Santaeulàlia-Llopis

European University Institute
Department of Economics

Education, HIV Status and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter?

Daniela Iorio and Raül Santaeulàlia-Llopis

EUI Working Paper **ECO** 2016/09

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Education, HIV Status and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter?*

Daniela Iorio
University of Bologna

Raül Santaeulàlia-Llopis
MOVE-UAB and Barcelona GSE

November 7, 2016

Abstract

We study the relationship between education and HIV status using nationally representative data from 39 Demographic and Health Surveys (DHS) in Sub-Saharan Africa. First, we construct an innovative algorithm that systematically defines aggregate stages of the HIV epidemic in a comparable manner across time and across space. Second, we exploit the variation in the aggregate HIV stages in the DHS data, and find that the education gradient in HIV shows a U-shaped (positive-zero-positive) pattern over the course of the epidemic. Further, educational disparities in the number of extramarital partners are largely consistent with the evolution of the education gradient in HIV.

*We thank Pascaline Dupas, Hans-Peter Kohler, Adriana Lleras-Muney, José-Víctor Ríos-Rull and Susan Watkins for helpful comments. We thank participants at the College of Medicine in Blantyre, the XXVI IUSSP International Population Conference (Marrakech) for very helpful comments and discussions. Daniela Iorio acknowledges financial support from the Barcelona GSE and the government of Catalonia. Raül Santaeulàlia-Llopis thanks the Weidenbaum Center at the Washington University in St. Louis for financial support. We also thank Alannah Glickman, Minho Kim, and Borek Vasicek for excellent research assistance.

1 Introduction

In Sub-Saharan Africa (SSA) most HIV infections are due to heterosexual intercourse, and risky sexual behavior is one of the most relevant margins that policy intervention can affect (Behrman and Kohler, 2012; Greenwood et al., 2013; DePaula et al., 2014; Nyqvist et al., 2015). If risky sexual behavior, and in turn HIV exposure, differs across education groups as the HIV epidemic evolves, then the timing of policy interventions targeted to specific educational groups is crucial for the effectiveness of these policies. Nowadays, however, the major international donors in the fight against HIV do not provide any guidance on specific targeting strategies across education groups.[1] This lack of policy advice could be explained by the fact that the current understanding of the sign and size of the relationship between education and HIV status lacks consensus (Beegle and de Walque, 2009). That is, the knowledge of which education groups are at major risk of being infected with HIV remains unclear to scholarship, with a large body of mixed evidence that we review below. We provide a potential explanation that could reconcile the mixed evidence, departing from the observation that the existing works have, almost invariably, used data from different aggregate stages of the HIV epidemic, while the education gradient in HIV may vary over the evolution of the epidemic.

That the HIV epidemic in SSA evolves differently across countries and that these countries are at different stages of the HIV epidemic at any point in time is practically self-evident. In particular, we observe that the peak of HIV prevalence, the year of the HIV peak, the time it takes each country to reach its own peak, and the pace at which each country moves away from its peak differs greatly across SSA countries and over time. Based on these considerations, we propose an innovative unified macro framework that consists of a two-dimensional normalization of the HIV epidemic. Our definition of the stages of the HIV epidemic are analogous to the definition of stages of economic development (Lucas, 2004; Herrendorf et al., 2014) or the stages of the demographic transition (Galor and Weil, 2000; Lee, 2003; Greenwood et al., 2005). In the context of HIV, our normalization adjusts for both the country-specific size of the epidemic (HIV prevalence rate) and the associated time paths of the epidemic (speed at which HIV epidemic

[1] According to Kates et al. (2011), $6.9 billion was given by donor governments to international AIDS assistance in 2010. The United States is the largest resource provider for the global fight against AIDS, and it channels its aid through the President's Emergency Plan for AIDS Relief (PEPFAR). Initiated by President George W. Bush for 2003-2008, PEPFAR has continued its activity under the mandate of President Barack Obama, who renewed the efforts for 5 years with few changes in policy implementation. However, they increased the amount of money—to about 50% in countries with a generalized epidemic— spent on preventing sexual transmission via abstinence, delay of age of first sexual intercourse, monogamy, fidelity, and reduction in the number of sex partners. More recently, new PEPFAR funds have been channeled to include "men who have sex with men, people who inject drugs, and sex workers", see the remarks of the Secretary of State John Kerry for the PEPFAR 10th annivarsary celebration at `http://www.state.gov/secretary/remarks/2013/06/210770.htm`. See also UNAIDS (2015).

reaches its peak and it moves away from it) in a comparable manner across countries. This way, our macro framework systematically defines aggregate stages of the HIV epidemic taking into account the large degree of cross-country heterogeneity in both the HIV prevalence rates and the speed at which the HIV epidemic evolves. We then use the heterogeneity in the stages of the HIV epidemic in 39 Demographic and Health Surveys (DHS) to document the stylized dynamic relationship between education and individual HIV status across the stages of the HIV epidemic.

Our main finding is that the education gradient in HIV follows a significant U-shaped (positive-zero-positive) pattern as the epidemic evolves. In particular, when individuals live in an economy that is at the early stages of the epidemic, the HIV-Education gradient is significantly positive and remarkably high: one additional year of education is associated with 1.12 percentage point increase in the probability of being HIV-positive. In other terms, completing five additional years of schooling doubles on average the likelihood of being infected (the HIV prevalence is 5.18% in our sample). Interestingly, the educational disparities in HIV gradually vanish as the epidemic progresses past these early stages, to then revert to a positive education gradient in HIV in the more advanced stages of the epidemic, where an additional five years of education result in a 2.40 percentage point increase in the probability of being HIV positive. When we control for country and year effects, a specification that is quite demanding because it exploits only variation witin each country, we also find a significant U-shaped pattern for the education gradient, although the size of the effects of education on HIV is smaller.

There are heterogeneous effects by gender. While women and men share the U-shaped pattern of the education gradient in HIV, its magnitude is larger for women than for men. This gap is largest at the early stages of HIV development, and tends to disappear later on. Regarding women, completing five additional years of schooling is associated with a 7.40% rise in the probability of being HIV positive at the early stages of the epidemic, that is, an increase twice as large as that of men, 3.80% per five years of education. Thereafter, the education gradient in HIV substantially declines until it vanishes for women who live in the middle of the HIV epidemic. At this stage, the decline in the gradient is even sharper for men, and changes its sign, reaching a -0.65% per five years of schooling. Interestingly, the gradient reverts to positive in the more advanced stages of the epidemic for both women and men, respectively, 2.70% and 1.75% per five years of education.

To gain a better understanding of the dynamic relationship between education and the probability of infection, we explore educational disparities in the actual risky sexual behavior. Remarkably, the pattern of the education gradient in HIV closely resembles the pattern of the educational disparities in risky sexual behavior. While more education is associated with more extramarital partners at early stages of the epidemic (0.19 per five more years of schooling for women and

0.15 for men), this relationship rapidly and significantly declines in mid stages of the epidemic (0.06 and 0.05 extramarital partners per five more years of schooling for, respectively, women and men). Interestingly, in the most advanced stages of the epidemic the relationship between education and the number of extramarital partners significantly increases (0.11 per five more years of schooling for women and 0.15 for men). The fact that the HIV-Education gradient closely follows education disparities in risky sexual behavior across stages of the epidemic points out the important role of educational disparities in determining the HIV incidence.[2] In this context, we posit a simple theoretical framework with endogenous risky sex choices depending on education that accounts for these facts.

The rest of our paper is organized as follows. In Section 2, we discuss the related literature. In Section 3, we describe our data. We document the heterogeneity of the epidemic across SSA countries and propose a unified macro framework to define the aggregate stages of the HIV epidemic in a comparable manner across time and space in Section 4. The estimates of the stylized evolution of the HIV-Education gradient across the aggregate stages of the epidemic are in Section 5. The potential mechanisms behind the evolution of the gradient are explored using a simple theoretical framework with risky sex choice that differs across education groups in Section 6. Section 7 concludes.

2 Related literature

We are not the first to investigate the HIV-Education gradient. A large number of epidemiological studies and small-scale studies examined socioeconomic disparities in HIV without reaching neither conclusive nor generalizable answers.[3] While the current mixed evidence is likely to reflect differences in methodology, sampling strategy, and measures of socioeconomic indicators and HIV status, this may not entirely explain the differing conclusions reached by previous studies, which

[2] The prevalence of HIV is determined by the newly infected individuals as well as by the survival probabilities of the individual being infected in the past. Therefore, the educational disparities in the HIV prevalence could be the result of educational disparities in the incidence of HIV and/or educational disparities in the survival probabilities. The DHS data allow us only to focus on individual attitudes towards risky sexual behavior that might increase the probability of being HIV infected. Later on, we discuss the potential role that educational disparities in survival rates might play in shaping the HIV-Education patterns as the HIV epidemic evolves.

[3] See the reviews of Hargreaves and Glynn (2002), Wojcicki (2005), Strauss and Thomas (2007) and Beegle and de Walque (2009). Regarding socioeconomic disparities in HIV, Lopman et al. (2007) find HIV prevalence is higher in poorer groups in Manicaland Province, Zimbabwe. In KwaZulu-Natal Province, South Africa, Bärnighausen et al. (2007) suggest the highest HIV prevalence is in the middle wealth class. In Limpopo Province, South Africa, Hargreaves et al. (2007) finds no relation between HIV prevalence and wealth. In an interesting comparative study across four cities, Glynn et al. (2002) finds that in two cities in Kenya and Zambia educational status and HIV were unrelated. However, in another city in Cameroon highly educated women were less likely to be infected with HIV while education and HIV status was unrelated for men. Finally, in another city in Benin highly educated men were less likely to be infected with HIV while the education and HIV status was unrelated for women.

for instance overlooked the large differences in the evolution of the HIV epidemic across SSA countries. In a review of epidemiological studies Gregson et al. (2001) conclude that there could be temporal dynamics of the influence of socio-economic development on rates of HIV transmission, and in particular that the greater vulnerability of individual with a high socioeconomic status may be a transient feature of the early stages of epidemics. The findings of de Walque (2007) point in the same direction in his analysis of the HIV-education gradient in rural Uganda between 1989 and 2001. Importantly, these studies did not expect a rebound in the education gradient in HIV that we document. Recently, using nationally representative data from the Demographic and Health Surveys (DHS) for five SSA countries, Fortson (2008) finds education has a positive association with HIV status: adults with primary school are one half more likely to be infected than adults with no schooling conditioning on age, sex and area of residence (urban/rural). Using the DHS wealth index, de Walque (2009) also finds that wealth displays a positive association with HIV status. Mishra et al. (2007) find similar results for eight DHS countries in SSA.

We contribute to this literature in two major respects. First, we uncover the evolution of the HIV epidemic using a unified macro framework that systematically defines the aggregate stages of the HIV epidemic in a comparable manner across countries. Our approach addresses similar data challenges as those faced when defining the stages of economic development (Lucas, 2004; Herrendorf et al., 2014) or the stages of the demographic transition (Galor and Weil, 2000; Lee, 2003; Greenwood et al., 2005). Second, in the context of this unified macro framework, and using repeated cross-sections of DHS surveys, our analysis exploits a rich variation in the aggregate stages of the HIV epidemic to document the stylized dynamic behavior of the HIV-Education gradient along the course of the epidemic, as well as the evolution of educational disparities in sexual responses. Our findings emphasize that while at early stages of the epidemic the HIV-Education gradient is large and positive (roughly three times larger than its stationary counterpart, à la Fortson (2008)), the gradient decreases to levels that are not significantly different from zero as the epidemic evolves; a macro cross-country decline that resembles the results by de Walque (2007) for several sites in rural Uganda. The time span of our data allows us to pick a rebound in the education gradient in HIV in mature epidemics, which we link to a positive change in risky sexual behavior among highly educated individuals. Interestingly, this U-shaped pattern is much more prominent for women than for men.

Due to the fact that heterosexual intercourse is the major mode of HIV infection (Behrman and Kohler, 2012), the relationship between HIV, risky sexual behavior, and HIV knowledge has been extensively studied. For example, information campaigns that improve the knowledge about HIV risk infection may induce people to adopt safer lifestyles. In this direction, de Walque (2007) documents substantial behavioral change in rural Uganda associated with the ABC campaign

(Abstinence, Be faithful, and use Condoms). In a somewhat opposite direction, Dinkelman et al. (2006) find little evidence for sexual behavior change associated with better knowledge on HIV prevention. More recently, in a randomized experiment conducted in primary schools in Kenya, Dupas (2011) shows that campaigns on the relative risk of infection by the age of sexual partner leads females to choose younger partners.[4] Further, inaccurate ex-ante beliefs about possible individual status implies that HIV testing may lead to behavioral responses (Boozer and Philipson, 2000). Randomizing monetary incentives to learn HIV status, Thornton (2008) finds that individuals who learn their HIV status are three times more likely to buy condoms (two months later) compared with those individuals who do not learn their HIV status. Note that the sign of the revision might matter as well. For example, if individuals revise their beliefs downward they may engage into riskier sexual behavior (Anglewicz and Kohler, 2009; Delavande and Kohler, 2012). However, the behavioral response may cease in two years or less (Thornton, 2012). More recently, DePaula et al. (2014) estimate a risky-sexual behavior model with beliefs about individual HIV status, and find that a downward (upward) revision in the belief of being HIV positive increases (decreases) risky sexual behavior, i.e., the chances of having an extramarital partner. An interesting additional layer of complexity is given by the individuals response to the HIV testing of others', as that can also change individuals' beliefs about the underlying HIV prevalence (Godlonton and Thornton, 2013). These authors explore this mechanism to find that (downward) revisions of HIV prevalence induce a decrease in condom use, but not in the number of multiple partnerships. Finally, Greenwood et al. (2013) propose an equilibrium model with endogenous formation of short versus long sexual partnerships and Bayesian updates on individual's HIV status (including their partners') to explore the relationship between individual HIV status and HIV prevalence.

Finally, our study is broadly related to the literature examining educational disparities in health outcomes (Lleras-Muney, 2004; Cutler et al., 2006; Cutler and Lleras-Muney, 2010, 2011). More recently, Cutler and Lleras-Muney (2014) examines this relationship in both developed and developing countries. Within this group of studies our work relates more closely to those that allow the relationship between education and health to be nonstationary. This is the case of the "fundamental cause" literature described in Cutler et al. (2006)[5] in which the diffusion of information on technological improvements is an argument used to explain the changes in the education gradient in health.[6]

[4] Albeit not using nationally representative data, a set of papers describes changes in sexual behavior across education groups over time or in response to a specific policy-driven change in the environment (e.g., Asiimwe-Okiror et al. (1997), UNAIDS (1997), Gregson et al. (2001), de Walque et al. (2005) and Stoneburner and Low-Beer (2004).

[5] See also Link and Phelan (1995) and Link et al. (1998).

[6] For example, the more rapid adoption by highly educated individuals of medical innovations and surgical

3 The Data

The core of our exercise consists of examining the relationship between education and HIV education over the stages of the HIV epidemic. To address this question it would be ideal to use nationally representative long panel data for several SSA countries starting in the pre-HIV era. Unfortunately, available nationally representative data are neither long nor panel. However, from a macroeconomic perspective, we show it is possible to construct a normalized path of the patterns of HIV infection by education groups over the stages of the epidemic for several SSA countries to recover stylized patterns between education and HIV. To do so, we combine two sources of data: (i) cross-sectional data from the DHS, and (ii) aggregate data from the most recent World Population Prospects (WPP) provided by the United Nations.

The Demographic and Health Surveys. The DHS are based on nationally representative samples and are available for a large set of SSA countries. We consider the full sample of SSA DHS surveys for which individual HIV testing has been conducted (and available as of July 2014): Burkina Faso (2003, 2010), Burundi (2010), Cameroon (2004, 2011), Congo (2007), Côte d'Ivoire (2005, 2011), Democratic Republic of Congo (2007), Ethiopia (2005, 2011), Gabon (2012), Ghana (2003), Guinea (2005, 2012), Kenya (2003, 2008), Lesotho (2004, 2009), Liberia (2007), Malawi (2004, 2010), Mali (2006), Mozambique (2009), Niger (2006), Rwanda (2005, 2010), Senegal (2005, 2010), Sierra Leone (2008), Swaziland (2006), Tanzania (2003, 2007, 2011), Uganda (2011), Zambia (2007) and Zimbabwe (2005, 2010), for a total of 25 DHS countries. For a number of countries the survey was conducted in two consecutive years, so we can exploit variation across 51 country-year pairs, which provide sufficient observational heterogeneity on the stages of the HIV epidemic to obtain reliable estimates of the evolution of the education gradient over these stages.

While the DHS are primarily health interviews, they also contain cross-sectional information on individual socioeconomic characteristics, knowledge on HIV, several measures of risky sexual behavior (e.g., number of extramarital relationships and condom use) and most importantly, a large proportion of adult respondents have been tested for HIV.[7] We use this cross-referenced

treatmens for heart disease may help to explain the widening of the mortality differentials by education groups in developed countries; see Feldman et al. (1989), Preston and Elo (1995), Goldman and Smith (2005), and Elo (2009). Aizer and Stroud (2010) also find that heterogeneous responses in the smoking behavior across education groups—highly educated individuals respond faster and stronger—to the first publicized report of the negative effects of smoking on health—the 1964 Surgeon General's report "Smoking and Health". de Walque (2004) shows similar evidence for slightly earlier periods (after 1950) when information about the implications of smoking on health started to diffuse.

[7]The proportion of respondents who did not take the HIV test is .318 in the original whole sample (.098 and .432 among men and women, respectively). However, we find that the association between the likelihood of taking the HIV test and the educational attainment is virtually zero in the DHS sample. Further, our evidence

individual information harmonically collected across SSA countries. Our whole sample consists of a total of 402,670 individuals, of which 56.5% are women. We choose to explore HIV infection risk by education groups separately for women and men. The average HIV prevalence is 6.1% for women and 4.1% for men (respectively, panel A and B, Table 1). There is a substantial degree of heterogeneity in HIV prevalence across countries. The Gini index for the HIV prevalence across these 39 DHS (25 countries) is 0.54 for women and 0.56 for men, with a range from 0.5 to 31.2 for women and from 0.4 to 19.7 for men. Note that there is a substantial HIV gender gap of 2% that is roughly half of the total HIV prevalence for men. Interestingly, as we show below, this gender gap in HIV prevalence evolves across the stages of the HIV epidemic. We restrict our attention to HIV-tested adults men and women 15-49 years old who reported their schooling achievement, which is on average 3.2 for women and 3.8 for men. The urban population is roughly one third for both women and men.

Several aspects make DHS datasets appealing for our exercise. An important advantage of these data is that they provide unambiguous individual measures of individual HIV status, education, knowledge on HIV, and risky sexual behavior in a comparable manner across SSA countries. First, regarding individual HIV status, the DHS provides a direct measure as individuals have blood testing for HIV, so we do not rely on indirect proxy for HIV obtained from other health outcomes or biomarkers. Second, regarding education variables, DHS collect data on education (i.e., number of years of schooling and maximum degree attained) and also provide an asset-based wealth index.[8] Our preferred choice for measuring education is years of schooling—perhaps the most commonly used measure for education in the previous literature. The reasoning for our choice of years of schooling—rather than the DHS wealth index— is that while wealth is influenced by subsequent negative health conditions (such as HIV), or other shocks that will potentially determine one's health status in adulthood, educational attainment is not because, typically, education is completed before individuals in our sample—adults between 15 and 49 years of age—enter adulthood. However, we cannot entirely rule out the fact that investments in education might respond to changes in life expectancy. Indeed, Fortson (2011) suggests a significant negative effect of HIV on investment in children in a model where agents explicitly consider mortality risk when making human capital decisions.

Finally, regarding risky sexual behavior, we focus on (i) the number of sex partners (i.e., the extensive margin) in the past 12 months other than spouses and (ii) condom use in last intercourse

suggests the DHS non-response bias for HIV testing is minimal. We find that statistics for age, schooling, and residence computed for the sample of HIV-tested adults resemble the analogous ones in the overall male sample. These results are available upon request. See also the discussions in Fortson (2008).

[8] Unfortunately, the DHS do not collect data on income, with very few exceptions that document wage earnings. This task is not easy as many SSA populations are mostly rural and a large proportion of these households' resources come from unsold agricultural production; see a discussion in De Magalhães and Santaeulàlia-Llopis (2015).

(i.e., the intensive margin, quality). The number of extramarital partners is on average 0.15 for women and 0.41 for men (panel A and B, Table 1). This is consistent with the population of women not having extramarital partners being larger than men's, respectively 87.4% and 73.7%. The frequency on condom use is 9.5% for women and 20.5% for men. That is, women not only report less extramarital partners than men, but also less condom use in the last sexual intercourse than men. These statements are consistent, as men's last sexual intercourse for casual sex is more common than women's. These figures are conditional on individuals being sexually active. Interestingly, the proportion of women not sexually active is smaller, 15.6%, than the one of men, 21.7%.[9]

The World Population Prospects. To uncover the evolution of the HIV epidemic we use the data from the 2015 revision of the World Population Prospects (WPP) constructed by the United Nations (Department of Economic and Social Affairs, Population Division). The WPP 2015 provides estimates of the HIV prevalence rates from 1980 until 2014 (at the country level), and their projections from 2014 onward for a large set of SSA countries. [10] The additional data on country-specific ART coverage used in our robustness exercise are also from the WPP.[11]

Finally, for all our SSA countries we use data on real output per capita from the Penn World Tables and data on agricultural share of output from the World Bank Development Indicators. We use these data in our empirical analysis to control for country-specific stages of aggregate economic development.

4 The Stages of the HIV Epidemic

This section describes the stages of the HIV epidemic. First, we discuss a set of challenges that we argue a useful definition of the stages of the HIV epidemic must address (Section 4.1). From our macroeconomic perspective these challenges arise from country differences in HIV prevalence across time (i.e., the evolution of the HIV epidemic within a country) and across space (i.e., heterogeneity of HIV prevalence across countries within a given period). Second, we provide an

[9] The Gini index for the number of extramarital partners across these 39 DHS (25 countries) is 0.37 for women and 0.30 for men, with a range from 0.01 to 0.56 for women and from 0.08 to 1.17 for men. The Gini index for the frequency of condom use in last sexual intercourse is 0.42 for women and 0.31 for men, with a range from 0.5% to 37.7% for women and from 4.2% to 49.3% for men (Table 1).

[10] The WPP data represent the official 2014 estimates of UNAIDS. Until 2006 the UNAIDS estimates relied mostly on data aggregations collected from antenatal clinics that overestimated prevalence levels. Since 2008, the UNAIDS data belong to a downward revision largely originated by the appearance of nationally representative surveys such as the DHS, and do not suffer from overestimation problems. Indeed, the HIV prevalence levels computed from our DHS samples and the HIV prevalence levels from WPP are very similar. See also a detailed discussion of these data in Bongaarts et al. (2008).

[11] Source: United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects. Unpublished Data - Special Tabulations. We thank Patrick Gerland for sharing these data.

algorithm that circumvents those challenges by normalizing the HIV epidemic in both dimensions, time and space (Section 4.2). Our definition is provided in Section 4.3.

4.1 Challenges for a Definition of the Stages of the HIV Epidemic

The evolution of the HIV epidemic is largely heterogeneous. To illustrate this, we show in Figure 1 the country-specific time path of the epidemic for a selected subsample of countries.[12] In addition, we provide country-specific statistics of the HIV epidemic for our entire sample of 39 DHS country-year surveys (Table 2). The following patterns arise across time and space.

4.1.1 HIV Prevalence Differences Across Space

While the HIV prevalence levels largely differs across countries (columns 1 and 2, Table 2),[13] two countries with the same HIV prevalence are not necessarily at the same epidemiological stage.

Remark 1. *The HIV prevalence alone is not sufficient to define the aggregate stage of the HIV epidemic.*

We show this argument with two straightforward counterexamples. Although Malawi in 1998 and Zimbabwe in 2010 both share the same HIV prevalence of 14.4, Malawi reaches this infection rate at its HIV peak while Zimbabwe reaches it only 13 years after its HIV peak (Figure 1). Indeed, in 2010 Zimbabwe's HIV prevalence is 55% of its HIV peak prevalence (column 6, Table 2). Another interesting counterexample arises from the comparison of the DHS observations of Zimbabwe and Lesotho. The DHS observation of Zimbabwe in 2005 delivers an HIV prevalence of 19.2%, lower than that of Lesotho in 2004, 23.4%. Looking at this statistic only, we would infer that Lesotho is at later stages of the epidemic than Zimbabwe. However, we actually know that Zimbabwe's HIV peak occurred at a higher level and earlier, 29.1% at 2009, than that of Lesotho, 23.8% at 2007, which suggests an opposite ordering over stages. That is, the ordering of DHS countries by HIV prevalence is a mere artifact of the years in which DHS were collected.

One step to address the problematic use of the absolute size of the HIV prevalence as a measure of the stages of the epidemic is to compute relative size of HIV prevalence dividing

[12] This subsample of DHS countries consists of Burkina Faso, Cameroon, Guinea, Lesotho, Malawi, Rwanda and Zimbabwe. This subsample serves expositional purposes only as it is useful to highlight the heterogeneity of the evolution of the HIV epidemic across countries as we describe next. Many other subsample choices would be equally useful.

[13] For example, in year 2010, the HIV prevalence (in percentages) is 1.1 in Burkina Faso, 5.1 in Cameroon, 1.9 in Guinea, 11.7 in Malawi, 3.1 in Rwanda and 18.0 in Zimbabwe. Across all SSA countries, the inequality in HIV prevalence remains high across time with a Gini coefficient of 0.63 in 1990, 0.56 in 2000 and 0.57 in 2010. The SSA set consists of 44 countries. Similar figures are attained with our sample of 25 DHS countries with Gini's coefficients of 0.55 in 1990, 0.51 in 2000 and 0.55 in 2010.

country-specific observations of HIV prevalence by their corresponding HIV peaks. However, this poses a new set of drawbacks because countries not only differ in the HIV peak level but also in the year of their HIV peak (columns 3 and 4, Table 2).

Remark 2. *The relative HIV prevalence alone is not sufficient to define the aggregate stage of the HIV epidemic.*

To see this, note, for example, that while the DHS observations of Guinea in 2005 and Ghana in 2003 share the same relative HIV prevalence of .93 (column 6, Table 2), Guinea attains that relative size 5 years before reaching its peak, and Ghana does so 3 years after reaching its peak (column 6, Table 2). This observation suggests that two countries can be at different stages of the epidemic despite having the same relative HIV prevalence. Another interesting example is the one posed by the DHS observations of Rwanda 2005 and Uganda 2011. Both surpassed their respective peaks, and have the same relative prevalence level of .52 (column 6, Table 2). However, it took Rwanda 11 years to move from its peak to this relative prevalence, while it took Uganda almost twice as much time, 20 years, to reach the same relative prevalence (column 5, Table 2). The fact that the transition away from the peak is slower in Uganda than in Rwanda is, in itself, a phenomenon to which we would like our definition of the stages of the epidemic to be invariant. Constructively, the arguments posed here against the sole use of the absolute (or the relative) HIV prevalence to define stages of the epidemic also suggest what we need to add to our definition of the stages to resolve the exposed problems: some properties of the time path of the HIV epidemic.

4.1.2 HIV Prevalence Differences Across Time

The time-path of HIV epidemic largely differs across countries. In particular, a large degree of heterogeneity exists for the HIV peak year across SSA countries (column 3, Table 2). In our DHS sample, the peak of the HIV year ranges from 1991 in Uganda to 2010 in Guinea, Lesotho and Swaziland. This leads to the following remark.

Remark 3. *Time (calendar year) alone is not sufficient to define the aggregate stage of the HIV epidemic.*

This remark states that two countries that suffer the HIV epidemic are not necessarily at the same epidemiological stage at the same calendar year. This is straightforward. Lesotho and Guinea reach the peak of their respective HIV epidemic in 2010, while Zimbabwe is at more advanced stage of its epidemic in 2010, precisely 13 years ahead of its HIV peak in 1997 (Figure 1).

One step to correct for the country-specific year of the HIV peak is to compute the relative time, i.e., calendar year minus year of HIV peak (column 5, Table 2). However, a large degree of heterogeneity exists for the speed by which SSA countries move to the respective HIV peaks and the speed by which SSA countries move away from their respective HIV peaks. This leads to the following remark.

Remark 4. *Relative time (calendar year minus year of HIV peak) alone is not sufficient to define the aggregate stage of the HIV epidemic.*

To see this, note that the DHS observations of Ethiopia in 2011 and Malawi in 2010 both share the same time distance with respect to their own HIV; in both cases 12 years have passed between the peak and the DHS data collection. However, in those 12 years Ethiopia has managed to decrease its relative HIV prevalence to 0.41 (column 6, Table 2), while Malawi has only managed to decrease its relative HIV prevalence to 0.75. Again, as we noted for remark 2, the fact that the transition away from the peak of Malawi is slower than that of Ethiopia is, in itself, a phenomenon to which we would like our definition of the stages of the epidemic to be invariant. The relative time does not suffice to define stages of the HIV epidemic.

To address these four remarks at once, we propose a two-dimensional (2D) algorithm that normalizes both the HIV prevalence level and time.

4.2 A Two-Dimensional Normalization of the Evolution of the Epidemic Across Time and Across Space

This section builds a 2D algorithm that, for all countries, normalizes the country-specific level and time path of the epidemic, thereby making the evolution of the HIV epidemic comparable across countries. Once the evolution of the epidemic is normalized for all countries, the position of each DHS dataset on its associated epidemiological stage readily follows.

Algorithm 1. [A Two-Dimensional Normalization of the Evolution of the HIV Epidemic] *Given the time series of the level of HIV prevalence of each i, we follow three steps to conduct a 2D normalization of the level and time path of the HIV epidemic:*

1. *Interpolate the country-specific time path of prevalence for each country i, $\{\lambda_{i,t}\}_{t_0}^{t_p}$, for $p+1$ interpolation points (years), where p is a positive integer. Then, interpolate the aggregate (across countries) prevalence path as $\lambda_t = \frac{\sum_i^n \lambda_{i,t}\mu_{i,t}}{\sum_i^n \mu_{i,t}}$, where n is the total of number of countries and $\mu_{i,t}$ is the population level of country i at period t. Denote the country-specific interpoland function as $s_i : t \to [0, \max_t \lambda_{i,t}]$, where $\max_t \lambda_{i,t} \in [0,1]$ and $s_i \in \mathcal{S}$, where $\mathcal{S}$ is the collection of functions that can be written as a linear combination*

of a set of n-known linearly independent basis functions ψ_j, $j = 1, ..., n$,

$$s_i(t) = \sum_{j=1}^{n} \theta_j \; \psi_j(t)$$

with n unknown θ_j coefficients. Denote the aggregate interpoland as $s(t)$ where $s(t)$ shares the same properties as the country-specific interpolands $s_i(t)$. Importantly, note that $\max_t s_i(t)$ is not necessarily identical across countries or to the aggregate $\max_t s(t)$.

2. *Level normalization*

   (a) *Compute the country-specific peak prevalence,*

   $$s_i(t_*^i) = \max_t s_i(t), \tag{1}$$

   where $t_*^i = \arg\max_t s_i(t)$ is the period country i reaches its peak, $s_i(t_*^i)$. Redo equation (1) to obtain the aggregate peak $s(t_*)$ and aggregate peak period, $t_* = \arg\max_t s(t)$.

   (b) *Normalize the country-specific and aggregate interpolands by their respective peak prevalence,*

   $$\widetilde{s}_i(t) = \frac{1}{s_i(t_*^i)} \; s_i(t) \quad \text{and} \quad \widetilde{s}(t) = \frac{1}{s(t_*)} \; s(t),$$

   where $\widetilde{s}_i, \widetilde{s} : t \to \Lambda = [0, 1]$ and $\arg\max_t s_i(t) = t_*^i = \arg\max_t \widetilde{s}_i(t)$. Note now that $\widetilde{s}_i(t_*^i) = \widetilde{s}(t_*) = 1 \; \forall i$.

3. *Time normalization*

   (a) *For $t_0^i < t^i \leq t_*^i$, normalize the time interval between the initial period for which data are available, $t_0^i = 1980$, and the country-specific peak period, t_*^i, by the time interval between the aggregate initial period, $t_0 = 1980$, and the aggregate peak period, t_*. To do so, we compute the constant of time proportionality for the pre-peak era,*

   $$\alpha_i^L = \frac{t_* - t_0}{t_*^i - t_0^i}.$$

   For $t^i > t_*^i$, normalize the time between the peak period, t_*^i, and the period t_γ^i in which country i reaches a given threshold $\gamma \in [0, 1]$, that is, $t_\gamma^i = \widetilde{s}_i^{-1}(\gamma)$, over the *analogous aggregate interval with t_* and $t_\gamma = \widetilde{s}^{-1}(\gamma)$,*

$$\alpha_i^R(\gamma) = \frac{t_\gamma - t_*}{t_\gamma^i - t_*^i}.$$

Here, note that t_γ^i and t_γ may not occur at an interpolation node but elsewhere along their respective interpoland.

(b) *Normalize the time input of the country-specific interpolands by α_i^L and α_i^R,*

$$\tau = \alpha_i^L(t - t_*^i) \quad \textit{for} \quad t \leq t_*^i \tag{2}$$

$$\tau = \alpha_i^R(t - t_*^i) \quad \textit{for} \quad t > t_*^i \tag{3}$$

where $\tau \in T$ are the normalized units of time. Operations (2) and (3) compress/stretch the interpoland[14] to ensure that for $\tau \leq \tau_*$ (before the peak) the number of normalized periods τ that it takes each country to move from τ_0 to the peak are the same across countries,

$$\widetilde{s}_i^{-1}(1) - \widetilde{s}_i^{-1}(0) = \widetilde{s}_j^{-1}(1) - \widetilde{s}_j^{-1}(0) = \widetilde{s}^{-1}(1) - \widetilde{s}^{-1}(0) \quad \forall\, i, j\ ,$$

and for $\tau > \tau_*$ (after the peak) the normalized periods τ that it takes each country to move from the peak to a threshold of prevalence γ is the the same across countries,

$$\widetilde{s}_i^{-1}(\gamma) - \widetilde{s}_i^{-1}(1) = \widetilde{s}_j^{-1}(\gamma) - \widetilde{s}_j^{-1}(1) = \widetilde{s}^{-1}(\gamma) - \widetilde{s}^{-1}(1) \quad \forall\, i, j.$$

This allows us to define the evolution of the epidemic for each country and the aggregate,

$$\widetilde{s}_i : \tau \to \Lambda \quad \textit{and} \quad \widetilde{s} : \tau \to \Lambda, \tag{4}$$

in the same—hence comparable—2D normalized space (T, Λ).

To implement the algorithm we need to make two choices: the shape of the basis functions, $\psi(\tau)$, and the prevalence threshold γ for the time normalization after the peak. First, we specify $\widetilde{s}(\tau)$ as a B-spline with cubic pieces and solve for the θ_j coefficients accordingly. Our choice of splines as interpolands obeys our desired manageability of the interpoland given the size of the Lagrangian interpolation problem poised by 71 (1980-2050) interpolation data points. Second, our choice of γ responds to balance between minimizing the use of projection of U.N. data for our set of DHS countries and maximizing the number of countries that have already surpassed the

[14] The interpoland s_i horizontally compresses when $\alpha_i^k > 1$ with $k = \{L, R\}$ and expands otherwise.

threshold γ at the time of DHS data collection. Our search for this balance suggests a value of $\gamma = .8$. This choice of γ implies that more than half of the countries in our dataset have already passed the threshold. Our results are robust to alternative choices of γ. To see this, note that the value of γ does not alter the ranking of countries across stages of the epidemic.

We next apply our algorithm to the SSA countries for which U.N. estimates and projections of the HIV prevalence time path are available.[15] The results are depicted in Figure 2. Panel (a) shows the HIV prevalence level across time for each and all countries. We highlight (in orange) the country-year observations for which DHS data with HIV testing results are available. This panel portrays all challenges discussed in Section 4.1. Panel (b) shows the result of the first normalization of our algorithm, the HIV prevalence level normalization, which tackles the issues associated with the absolute HIV prevalence, but does not correct for having HIV peaks at different calendar years for different countries. Panel (c) shows the result of the second normalization of our algorithm, the time normalization, which forces all countries to peak at the same calendar year but does not resolve the issues associated with the absolute HIV prevalence level. Finally, panel (d) closes our algorithm by jointly applying the level and the time normalizations. These 2D normalization resolves the full set of challenges posed in Section 4.1 as the relative HIV prevalence peaks in all countries at the same period. As it is obvious from panel (d), after the 2D normalization the level and time path of the epidemic are entirely comparable across countries.

4.3 A Definition of the Stages of the HIV Epidemic

Note that the position of each country i on its normalized HIV time path at the period t_{DHS} at which its respective DHS data were collected can be easily computed by solving for τ_i in

$$\widetilde{s}_i^{-1}\left(\frac{\lambda_{i,t_{DHS}}}{s_i(t_*^i)}\right) = \tau_i.$$

Then, the stage of the HIV epidemic is the continuous real variable,

$$\omega(\tau,\zeta) = \frac{\zeta}{\tau - \tau_*} \to \mathcal{R}^1, \tag{5}$$

with the pair (τ,ζ) belonging to the 2D normalized space, $T\times\Lambda$. Geometrically, $\omega(\tau,\zeta)$ represents the slope of the arrays from the origin in the (T,Λ) space with the following limiting properties:

$$\lim_{\tau\to\tau_*^-}\omega(\tau,\zeta) = -\infty,\ \lim_{\tau\to\tau_*^+}\omega(\tau,\zeta) = \infty,\ \text{ and } \lim_{\tau\to-\infty}\omega(\tau,\zeta) = \lim_{\tau\to+\infty}\omega(\tau,\zeta) = 0.$$

[15] The interpolation Lagrangian points, that is, the country-specific prevalence time series, $\lambda_{i,t}$, are retrieved from the U.N. population division estimates and projections until 2050 (medium-variant).

To conduct our empirical exercise, we discretize the continuous variable that defines the epidemiological stages in (5).[16]

Definition 1. [Stages of the HIV Epidemic] *Given a set of stage thresholds* $\{\zeta_0, ..., \zeta_j, ..., \zeta_n\}$ *with* $\zeta_j > \zeta_{j+1}$ *for all* j*, the stage* j *of the HIV epidemic consists of all pairs* $(\tau, \zeta) \in T \times \Lambda$ *such that* $\omega(\widetilde{s}^{-1}(\zeta_{j+1}), \zeta_{j+1}) \leq \omega(\tau, \zeta) \leq \omega(\widetilde{s}^{-1}(\zeta_j), \zeta_j)$*, where* $\widetilde{s}(\tau)$ *is the normalized (population-weighted) aggregate of the HIV epidemic defined in (4).*

Our choice for the stage thresholds $\{\zeta_0, ..., \zeta_j, ..., \zeta_n\}$ pursues the maximization of both countries per stage of the epidemic and number of stages. To do so, we set $\zeta_0 = 1$ and $\zeta_j = \zeta_1 - .05j \; \forall j$. The results of this exercise are shown in Figure 3 where each data point $(\tau_i, \widetilde{s}_i)$ represents a DHS dataset. This implies the following allocation of DHS datasets, $\omega(\tau_i, \widetilde{s}_i(\tau_i))$, over stages of the epidemic as follows:

- Stage ≤0: Cameroon 2004, Guinea 2005, Guinea 2012, Lesotho 2004/05, Lesotho 2009/10, Mozambique 2009, Senegal 2005, Sierra Leone 2008, Swaziland 2006/07.
- Stage 1: Cameroon 2011, Cote d'Ivoire 2005, Democratic Republic Congo 2007, Ethiopia 2005, Ghana 2003, Kenya 2003, Liberia 2006/07, Malawi 2004/05, Mali 2006, Niger 2006, Senegal 2010/11, Tanzania 2003/04, and Zambia 2007.
- Stage 2: Gabon 2012, Malawi 2010, Tanzania 2007/08, Zimbabwe 2005.
- Stage 3: Congo Brazaville 2009, Cote d'Ivoire 2011, Kenya 2008/09, Rwanda 2005, Tanzania 2011/12, Zimbabwe 2006.
- Stage ≥4: Burkina Faso 2003, Burkina Faso 2010, Burundi 2010, Cote d'Ivoire 2012, Ethiopia 2011, Niger 2012, Rwanda 2010/11, Uganda 2011 and Zimbabwe 2010/11.

Figure 3 depicts these allocations in two dimensions in the normalized space $T \times \Lambda$. It is identical to panel (d) in Figure 2 with the addition of the arrays initiating from the origin that define the breakdown of discrete stages of the HIV epidemic defined above. It is relevant to note that the DHS observations covers the entire evolution of the HIV epidemic except its initial rise, which is not surprising given that the first DHS surveys with HIV testing were conducted in 2003. In any case, the DHS observations provide a large degree of heterogeneity across country

[16] To capture the evolution of the HIV-Education gradient over the HIV epidemic we need to interact education with either some specific function (e.g., a quadratic or cubic polynomial) of $\omega(\tau, \zeta)$ or with a discretized version of $\omega(\tau, \zeta)$. We prefer to follow the latter approach and partition the continuous variable $\omega(\tau, \zeta)$ into the discrete pieces which delivers a cleaner interpretation of the estimates of the HIV-Education gradient across stages.

positions over the normalized HIV epidemic. The large heterogeneity across stages of the HIV epidemic is sufficient to provide reliable nonstationary estimates of the HIV-Education gradient.[17]

In Section 3, we have highlighted the presence of a gender gap defined as HIV prevalence of women minus HIV prevalence of men (ages 15-49). We are now in the position to assess the evolution of this gender gap across stages of the HIV epidemic, as depicted in Figure 4. Before the peak of the epidemic ($\tau = 0$ in Figure 4) is reached, the HIV gender gap is negative with men having on average 0.5% more HIV prevalence than women. Interestingly, the gender gap disappears at the peak of the epidemic, and turns positive and increasing after the peak with women having on average 0.5% more HIV prevalence than men (0.8% in the more advanced stages of the epidemic). While we use the epidemiological stage jointly defined for both women and men throughout our empirical analysis, we can also implement our algorithm for women and men separately. When doing so, we find that men tend to lead the epidemic by reaching their HIV peak earlier than women, but also by progressing away from it at a faster rate as well (see our Appendix Figures A-1 and A-2).

4.4 Discussion

The normalization procedure that we have proposed to define stages of the HIV epidemic is drawn upon ideas developed by macroeconomists to define the stages of the demographic transition and the stages of economic development.

As it is the case with the HIV epidemic, the demographic changes behind the demographic transition occur at different calendar years and at different speeds for different countries. To take this into account, the demographic model controls for calendar year and speed of stage completion with a time-normalization analogous to the one we propose, and in turn defines the demographic stages in a similar way to ours. Precisely, many countries have gone (or as still going) through a first demographic stage of high mortality and fertility, followed by a second demographic stage in which mortality drops and, with some period lag, fertility declines, and finally a third demographic stage of low mortality and low fertility (Lee, 2003). This description of the stylized joint behavior of mortality and fertility exemplifies how the calendar year, the speed at which each country completes each demographic stage, and the country-specific levels of the variables of interest (mortality and fertility) needs to be controlled for to recover stylized facts across stages of the demographic transition.

With regard to the aggregate stages of economic development (or structural transformation

[17] For the country of Niger and Senegal, we lack of projected paths, and therefore we cannot implement the time-normalization. We allocate the DHS surveys of these countries across stages by visual inspection. We find reassuring that the exclusion of Niger and Senegal from our analysis does not alter our results.

out from agriculture), the heterogeneity across time and across space depends on the fact that countries take off at different calendar years and move across stages of economic development at different speeds (Hansen and Prescott, 2002; Gollin et al., 2002, 2007). For example, China's income per capita raised by a factor of 6 between 1989 and 2009, while it took the U.K. from 1820 to 1970 to generate such growth. That is, China has grown at roughly 7.5 times the speed of the first industrial revolution (Bolt and van Zanden, 2013). With a definition of the stages of development that nets out calendar years and speed, we can study stylized economic relationships along the process of economic development such as the skill wage premium (Buera et al., 2015).

In sum, the spirit of our exercise is similar to what has been done to describe the stages of demographic transitions and the stages of economic development. First, we provide an algorithm that defines aggregate stages of the HIV epidemic netting out the calendar year effects, the country-specific speeds of the HIV epidemic, and the HIV prevalence levels. Second, we study the stylized dynamic relationship between HIV status and education across stages of the HIV epidemic.

5 The HIV-Education Gradient

Our empirical analysis consists of posing a simple econometric specification suitable for documenting the potentially nonstationary behavior of the HIV-Education gradient.

5.1 Econometric Specification

We consider a linear probability model (LPM) where the HIV-Education gradient is allowed to change over the stages of the HIV epidemic (j) defined in Section 4.3. Since for a large set of countries there are at least two cross-sections in our sample, and individuals in different periods are not the same people for each country (g), we index the variables by a double subscript. Namely, $i(t) \in \{ 1, ..., N_t\}$ denotes the individuals in cross-section t. Let $s_{i(t),t,j}$ denote the educational attainment of an individual $i(t)$ that lives in stage j of the HIV epidemic at time t. Also, let $y_{i(t),t,j}$ be the individual's HIV status, a dummy variable equal to one if the individual HIV testing result is positive and zero otherwise. We estimate a linear projection of the type,

$$y_{i(t),t,j} = \alpha_0 + \sum_{j>0} \alpha_j \mathbf{1}_j + \left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j \right) s_{i(t),t,j} + \beta \; x_{i(t),t} + \psi \; m_{g,t} + \theta_t \mathbf{1}_t + \theta_g \mathbf{1}_g + \varepsilon_{i(t),t}, \quad (6)$$

where $\mathbf{1}_j$ is an indicator function that is equal to one when the stage of the HIV epidemic is j and zero otherwise. That is, if the stage of the epidemic is $j = 0$ then the intercept is α_0 and the slope is γ_0. However, for each the stage of the epidemic is $j > 0$, the associated intercept is $(\alpha_0 + \alpha_j)$

and the slope is $(\gamma_0 + \gamma_j)$. Namely, γ_j is the difference in the HIV-Education gradient between individuals that are in stage j and stage 0 of the epidemic. This implies that the HIV-Education gradient is $\gamma_0 + \gamma_j$ for each epidemiological stage j. We cluster the individual observations at the country level to account for any unobserved shock that correlates observations within a country. Given that the number of countries is 25, we use the wild cluster bootstrap from Cameron et al. (2008) to get better approximations to asymptotically valid standard errors.[18]

The vector $x_{i(t),t}$ corresponds to a set of individual characteristics that are likely to be correlated with both education and HIV status. Hence, controlling for these characteristics reduces the impact of omitted variables bias. Precisely, we find it is important to control for the type of area in which agents live because the HIV prevalence is, on average, higher in urban than in rural areas (respectively, 6.73% and 4.41% in our whole sample), and it is in urban areas where adult education levels are also higher (the average number of years of schooling in rural area is 3.23, while in urban areas reaches 3.83). Thus, a positive association between education and HIV may be driven by the fact that people living in urban areas are both more likely to be HIV-positive and more educated. Similarly, we also control for age because HIV prevalence is increasing with age (the DHS age sample is 15-49), and education is negatively correlated with age as younger cohorts are more educated than older cohorts.

We also control for time varying country-specific economic variables, $m_{g,t}$, which correct for the stage of economic development in which each country is. To do so, we use measures of output per capita and share of agricultural output following the literature on structural transformation. We also include year dummies (θ_t) and country dummies (θ_g) to pick up any spurious correlation between the regressors and the dependent variable. To the extent that such contextual effects affect all individuals in a country in a similar manner, the country dummies will sweep them up. All our specifications are weighted least squares regressions, where the weights are proportional to the relative population size of each country. By doing so, when we pool a number of countries in the same stage of the HIV epidemic, the relative DHS sample size of a given country corresponds to the relative population size of the country. We then combine these weights with the individual weights provided by the DHS surveys.

[18] The results are robust to clustering at the country-year level. While we believe it would be interesting to explore also the within-country variation (e.g., across regions), it is not feasible to recover the epidemiological stages using our algorithm proposed in section 4 due to data limitations about the evolution of the HIV epidemic at the regional level; recall that the algorithm would require complete time-series of HIV prevalence for each region within a country and these estimates are generally provided at the national level (UNAIDS, 2015).

5.2 Results

We first consider a linear probability model where the dependent variable is the individual HIV status. The estimates of the HIV-Education gradient using the whole sample are reported in Table 3. We then re-conduct our analysis separately for women and men in Table 5.

5.2.1 Stationary Specification

To study the stationary gradient we restrict the econometric model (6) with $\alpha_j = \gamma_j = 0$ for all $j > 0$. We find that the stationary HIV-Education gradient is highly significant and positive (column 1, Table 3). The probability of being HIV infected increases by 0.43% per year of schooling. This suggests that completing five additional years of schooling increases the probability of being HIV positive by 2.15%, which is not small if we consider that the HIV prevalence is 5.18% in our sample.[19] Further, the probability of being HIV positive is higher for women (by 2.24%), for urban areas (by 2.12%), and it increases significantly with age (0.25% per year of age).[20] Aggregate variables denoting the stage of development such as the agricultural share of output and output per capita are negatively related with the probability of being infected. Next we explore how much this gradient changes as the HIV epidemic evolves.

5.2.2 Non-Stationary Specification

Our non-stationary specification follows the econometric model in (6). Our key finding is that the HIV-Education gradient is significantly nonstationary and displays a positive-zero-positive U-shaped pattern over the stages of the HIV epidemic.

Focusing on our benchmark specification (column 2, Table 3), we find that at Stage 0 an additional schooling year raises the probability of being infected by γ_0 = 1.12%. That is, for individuals in an economy that is at early stages of the epidemic the HIV-Education gradient is significantly positive and remarkably high (roughly three times larger than that of the stationary specification). Interestingly, as the HIV epidemic evolves, the HIV-Education gradient rapidly declines. At Stage 1 the rise in the probability of being infected associated with one additional year of schooling is $\gamma_0 + \gamma_1$ = 0.51%, i.e., less than one-half of its value at Stage 0, and it is significantly different from zero at 1% level (column 2, panel A of Table 4). The educational disparities in HIV then vanish as the epidemic reaches Stage 2, where we cannot reject the null that $\gamma_0 + \gamma_2$ = -.04% is different from 0 (column 2, panel A in Table 4). As we move away

[19]These findings are consistent with those obtained by Fortson (2008), who specifies a similar stationary econometric model for five DHS countries.

[20]While we introduce age linearly, we do find that the estimated coefficients for the HIV-Education gradient are robust when age enters non-linearly.

from Stage 2, the HIV-Education gradient becomes increasingly positive as the epidemic evolves with $\gamma_0 + \gamma_3 =$ 0.19% and $\gamma_0 + \gamma_4 =$ 0.48% in Stages 3 and 4, respectively. This way, the HIV-Education gradient bounces back reaching a significant gradient in Stage 4 that is almost half the size of the gradient in Stage 0. Note that both the initial decline of the HIV-Education gradient from Stage 0 to Stage 2 and its posterior rebound from Stage 2 to Stage 4 are both significant. The size of the rebound from Stage 2 to Stage 4 is a significant 0.52% (column 2, panel B in Table 4). We conclude that the HIV-Education gradient exhibits a positive-zero-positive U-shape pattern over stages of the HIV epidemic. To illustrate this pattern, Figure 5 shows the isomorphic representation of the estimated HIV-Educ gradient across stages of the epidemic, $\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j$ for each j (panel A, Table 4).[21] Similar results are attained if we consider only the sexually active subsample (Appendix Table A-1), with an HIV-Education gradient of 1.23% at Stage 0, 0.59% at Stage 1, -0.06% at Stage 3, 0.22% at Stage 3 and 0.56% at Stage 4.

The U-shape pattern of the HIV-Education gradient across stages of the HIV epidemic is robust to the addition of year dummies, country dummies, and year and country dummies (columns 3 to 5 in Table 3). When we control for year dummies, the HIV-Education gradient is significantly different from zero for all stages except Stage 2 showing the same positive-zero-positive behavior as our benchmark (column 3 in Table 4). When we control for country dummies the magnitude of the HIV-Education gradient is smaller (columns 4 and 5, panel A, Table 4) but the U-shape pattern is preserved and remains significant, that is, γ_2 and γ_4 are significantly different from each other (column 4 and 5, panel B, Table 4). This way, while the specification with country-fixed effects is quite demanding, as it exploits only variation witin each country, the direction of the results discussed above remains unchanged. In particular, the inital decline and posterior rebound remain significant. Finally, the same U-shape pattern arises in panel B of Table 3 where we re-estimate the model for each stage separately.

Given that women are significantly more likely to be HIV positive (an estimate of 2.24% across specifications, Table 3), and that this relationship depends on the stage of the epidemic (Section 4), we also investigate whether the HIV-Education gradient differs across gender. Our results for women and men are, respectively, in panel A and B of Table 5. The stationary HIV-Education gradient is 0.48% per schooling year for women and 0.33% for men. We also show the nonstationary behavior of the HIV-Education gradient for women and men across stages of the epidemic in Figure 6. The main result is clear. The HIV-Education gradient is significantly U-shaped over the stages of the HIV epidemic for both women and men, but the estimates are larger for the former group. This gap is apparent in the early stages of HIV development, and

[21] Our results also hold under a Probit specification. The partial effects (and p-values) are as follows: $\gamma_0 =$ 0.691% (0.000), $\gamma_0 + \gamma_1 =$0.465% (0.000), $\gamma_0 + \gamma_2 =$0.018% (0.674), $\gamma_0 + \gamma_3 =$0.135% (0.133), and $\gamma_0 + \gamma_4 =$0.339% (0.004). Regarding the rebound, we also reject the null that $\gamma_4 - \gamma_2 =$0 (0.018).

tends to disappear later on. Regarding women, an additional year of education is associated with a 1.48% rise in the probability of being HIV positive at Stage 0, that is, an increase twice as large as that of men, 0.76% per year of education. Thereafter, the education gradient in HIV substantially declines until it vanishes for women who live in Stage 2 of the HIV epidemic. At this stage, the decline in the gradient is even sharper for men, and changes its sign, reaching a -0.13% decline in the probability of being HIV positive per year of schooling. Interestingly, the gradient reverts to positive in the more advanced stages of the epidemic for both women and men: about 2.7% and 1.75% per five years of education, respectively. in Stage 4. Regarding the significance of $\gamma_0 + \gamma_j$, the women HIV-Education gradient is significantly positive (at 1% level) except for stage 2 and 3, and the men HIV-Education gradient is significant for each stage j, which includes the negative—but quantitatively minor—gradient at stage 2. The rebound of the HIV-Education gradient in the most advanced stages of the epidemic, i.e., $\gamma_4 - \gamma_2$, is significant for both women and men at 1% level across all specifications.

To summarize, the HIV-education gradient shows a U-shape positive-zero-positive pattern across stages of the epidemic. This stylized fact is twice more sizable for women than for men at early stages of the epidemic, the gradients across genders tend to equalize in later stages.

6 What Is Behind the Evolution of the HIV-Education Gradient?

Note that estimates based on HIV prevalence reflect risk factors for survival as well as those for incidence. We begin by considering those factors that may affect the incidence rate. In SSA, where the majority of the HIV infections occur through sexual intercourse (Behrman and Kohler, 2012; Greenwood et al., 2013), a natural candidate to account for the nonstationary behavior of the HIV-Education gradient is differences in risky sexual behavior across education groups.

We next consider a simple model of risky sex to illustrate how educational disparities in HIV may change as the epidemic evolves over time.

6.1 A Theoretical Interpretation

We posit a simple but flexible life cycle model where individuals differ in education, i.e., schooling years s. Agents live for two periods young and old. When agents are young they are HIV$-$, and the probability of being HIV$+$ in old age depends on risky sexual behavior choice. Agents maximize their lifetime utility choosing over consumption and risky sex,

$$\max_{\{c_0 \geq 0, c_1 \geq 0, x \geq 0\}} u(c_0, x) + \mathbb{E}\left(\lambda(x)\gamma^+(s) + (1-\lambda(x))\gamma^-(s)\right) u(c_1)$$

subject to $c_0 + px = y_0(s)$ and $c_1 = y_1(s)$, where c_0 is consumption today, c_1 is consumption tomorrow, and income y is strictly increasing and strictly concave with respect to education, which is exogenously given.[22] Risky sexual activities, $x \in R^+$, are conducted only when young and priced at p. More x can be interpreted as having sex with more partners or as having sex without condom. The presence of safe sex, which we omit, is innocuous for our exposition.[23] The probability of infection, $\lambda(x)$, is strictly increasing and strictly concave with risky sex, i.e., $\lambda_x > 0$ and $\lambda_{xx} < 0$. The survival probability depends on HIV status with $\gamma^+(s) < \gamma^-(s)$. We also assume that survival probabilities increase with education.

The first order condition with respect to x is,

$$u_x = pu_c + \beta\lambda_x \left(\gamma^-(s) - \gamma^+(s)\right) u(y_1(s)).$$

We assume that $u(c_0, x) = \ln c_0 + \kappa x$. This way, $u_x = \kappa$ is the marginal benefit of one extra unit of risky sex. An additional unit of risky sex has a marginal cost today and a marginal cost tomorrow. Today's marginal cost is $pu_c = \frac{p}{c}$, i.e., the forgone current consumption. Tomorrow's marginal cost, $W(x, s) = \beta\lambda_x (\gamma^-(s) - \gamma^+(s)) \ln(y_1(s))$, is the loss of continuation value for one unit of risky sex today and incorporates the fact that more risky sex today increases (decreases) the probability of infection (survival). The optimal amount of sex occurs when the marginal benefit equates the total marginal cost (panel (a), Figure 7).

Solving for the first order conditions, optimal risky sex is,

$$x = \frac{y_0(s)}{p} - \frac{1}{\kappa - W(x, s)}, \tag{7}$$

The key aspect of this model is that the infection rate is endogenous to s; that is, for a given s, agents optimally choose x which in turn determines $\lambda(x)$. We denote this dependence as λ_s.

Note that the optimal amount of risky sex is strictly increasing in education when risky sex has no future consequences, i.e., tomorrow's marginal cost is zero. When agents take into account future costs of risky sex, the optimal amount of sex will be smaller (see also Panel (a) of Figure 7, where we plot the marginal benefit and costs of risky sex for a given set of parameters and functional forms. Further, education affects risky sex through two channels going in opposite directions: i) as above, more education decreases today's marginal cost per unit of sex due to higher current income, which unambiguously increases the amount of sex—sex is not an inferior

[22] The assumption that education is positively related to income has been used in other studies aimed at explaining why education might be related to health, see e.g. Cutler et al. (2006)

[23] For example, we can add a third input to the utility function representing safe sex which can be priced at zero and individuals always satiate.

good; ii) more education increases tomorrow's marginal cost per unit of sex due to higher future income and higher chances of survival. We can show that the sign of the comparative static remains positive under reasonable assumptions though: if $(\gamma^- - \gamma^+)$ is sufficiently decreasing in s, and if λ_{xx} is relatively small. A graphical description of the qualitative effects of education on risky sexual behavior is in panel (b) and(c) of Figure 7.

A number of considerations follow. First, this simple theoretical model accounts for the fact that the education gradient in risky sex is positive. Second, agents may tend to disregard the future consequences of being HIV infected when the HIV epidemic takes off, that is when they are not aware of the harmful side effects of HIV. As a result, the education gradient in risky sex is high and positive at early epidemics. As the HIV progresses and agents realize about the true costs of risky sex, the trade-off highlighted above leads this gradient to decrease, and may eventually vanish. Third, the rebound in the gradient may occur as the gap in the survival probabilities $(\gamma^- - \gamma^+)$ shrink faster for the highly educated than for the low educated individuals. This is in line with the introduction of antiretroviral therapy (ART). A parsimonious extension of our model to incorporate ART suggests that ART can help explain the increases in risky sexual behavior and hence the rebound in the HIV-Education gradient if the more educated have more access to ART or as long as access is costly (see Section 6.2).

Last but not least, the fact that more education increases risky sex implies that the more educated should bear more HIV infections. It is straightforward to see that this theoretical result delivered by the two period model carries across stages when we allow overlapping generations. This way, the model can capture the U-shape dynamics between education and HIV status along the HIV epidemic if educational disparities in risky sexual behavior also displays that U-shape across stages of the epidemic. To see this, we need to keep track of the endogenous composition of HIV individuals derived from our model. A simple population projection matrix model allows us to do so. Denote by $\mu^h_{s,\tau}$ the measure of individuals with s years of schooling and HIV status $h \in \{-, +\}$ in stage of the epidemic τ. Then, the evolution of a given education group s over epidemiological stages can be traced by

$$\begin{bmatrix} \mu^+_{s,\tau} \\ \mu^-_{s,\tau} \end{bmatrix} = \begin{bmatrix} \gamma^+_s & \lambda_s \, \gamma^-_s \\ 0 & (1-\lambda_s) \, \gamma^-_s \end{bmatrix} \begin{bmatrix} \mu^+_{s,\tau-1} \\ \mu^-_{s,\tau-1} \end{bmatrix} + \begin{bmatrix} \phi^+_{s,\tau-1} \\ \phi^-_{s,\tau-1} \end{bmatrix}, \tag{8}$$

where $\gamma_s = \{\gamma^+_s, \gamma^-_s\}$ are the survival rates of individuals with s schooling years and these differ by HIV status, λ_s is the rate at which individuals with s schooling years become infected with HIV, an equilibrium object that depends on the choice of risky sex (7), and $\phi_{s,\tau-1} = \phi^-_{s,\tau-1} + \phi^+_{s,\tau-1}$ denotes an exogenous new adult population—individuals that become adults at stage τ with schooling years s.

Using the population law of motion (8), the probability of being HIV-positive for individuals in education group s in stage τ (i.e., the HIV prevalence of schooling group s in stage τ) is

$$HIV_{s,\tau} = \frac{\mu^{+}_{s,\tau}}{\mu^{+}_{s,\tau} \;+\; \mu^{-}_{s,\tau}} = \frac{\gamma^{+}_{s}\;\mu^{+}_{s,\tau-1} \;+\; \lambda_s\,\gamma^{-}_{s}\;\mu^{-}_{s,\tau-1} \;+\; \phi^{+}_{s,\tau-1}}{\gamma^{+}_{s}\;\mu^{+}_{s,\tau-1} \;+\; \gamma^{-}_{s}\;\mu^{-}_{s,\tau-1} \;+\; \phi^{+}_{s,\tau-1} \;+\; \phi^{-}_{s,\tau-1}},$$

and we compute the HIV-Education gradient between education groups s and $s+1$ at τ as[24]

$$G(\mu_{s+1,\tau},\mu_{s,\tau})(\lambda_{s+1},\lambda_s,\gamma_{s+1},\gamma_s,\phi_{s+1,\tau},\phi_{s,\tau}) = \frac{HIV_{s+1,\tau} - HIV_{s,\tau}}{HIV_{s,\tau}}.$$

We are interested in which factors can generate changes in the HIV-Education gradient over epidemiological stages. After some algebra, the following implications arise for the gradient:

$$\frac{\partial G}{\partial \lambda_{s+1}} > 0,\ \frac{\partial G}{\partial \lambda_{s}} < 0,\ \frac{\partial G}{\partial \gamma^{+}_{s+1}} > 0,\ \frac{\partial G}{\partial \gamma^{+}_{s}} < 0,\ \frac{\partial G}{\partial \phi^{+}_{s+1}} > 0,\ \text{ and }\ \frac{\partial G}{\partial \phi^{+}_{s}} < 0.$$

This summarizes the two main channels affecting the HIV-Education gradient. First, an increase in the rate of new HIV infections of the more educated (i.e., λ_{s+1}) relative to the less educated (i.e., λ_s) increases the HIV-Education gradient. This holds theoretically because $\lambda_{s+1} > \lambda_s$ from (7), that is, a higher s is associated with higher risky sexual activity x. This way, to validate this prediction we need to check whether a Risky Sex-Education gradient tracks the HIV-Education gradient across epidemiological stages. Second, within the HIV-positive population, an increase (decrease) in the survival rate of the more educated (i.e., γ^{+}_{s+1}) relative to the less educated (i.e., γ^{+}_{s}) implies an increase (decrease) in the HIV-Education gradient. As per our model (7), another interesting aspect of increasing γ^{+} is that this implies an increase in risky sexual activity which, in turn, increases the probability of infection.

Last, we note that increases (decreases) in the number of more-educated individuals in the education composition of HIV-positive individuals who become adults (i.e., ϕ^{+}_{s+1}) relative to the number of less educated (i.e., ϕ^{+}_{s}) implies increases (decreases) in the HIV-Education gradient.[25]

[24] Precisely, one additional year of schooling is associated with a change in the probability of HIV infection by

$$G(\mu_{s+1,\tau},\mu_{s,\tau})(\lambda_{s+1},\lambda_s,\gamma_{s+1},\gamma_s,\phi_{s+1,\tau},\phi_{s,\tau}) = \frac{\dfrac{\gamma^{+}_{s+1}\;\mu^{+}_{s+1,\tau-1} + \lambda_s{+}1\;\gamma^{-}_{s+1}\;\mu^{-}_{s+1,\tau-1} + \phi^{+}_{s+1,\tau-1}}{\gamma^{+}_{s+1}\;\mu^{+}_{s+1,\tau-1} + \gamma^{-}_{s+1}\;\mu^{-}_{s+1,\tau-1} + \phi^{+}_{s+1,\tau-1} + \phi^{-}_{s+1,\tau-1}}}{\dfrac{\gamma^{+}_{s}\;\mu^{+}_{s,\tau-1} + \lambda_s\;\gamma^{-}_{s}\;\mu^{-}_{s,\tau-1} + \phi^{+}_{s,\tau-1}}{\gamma^{+}_{s}\;\mu^{+}_{s,\tau-1} + \gamma^{-}_{s}\;\mu^{-}_{s,\tau-1} + \phi^{+}_{s,\tau-1} + \phi^{-}_{s,\tau-1}}} - 1,$$

and note that $\gamma^{-}_{s}\mu^{-}_{s,\tau-1} + \phi^{+}_{s,\tau-1} + \phi^{-}_{s,\tau-1} > \lambda_s\gamma^{-}_{s}\mu^{-}_{s,\tau-1} + \phi^{+}_{s,\tau-1}$.

[25] Mother-to-child HIV transmission may enter through ϕ, although this margin represents a minor source of aggregate HIV prevalence (UNAIDS, 2015). Also, a decline in the number of HIV-positive individuals who reach adulthood with more years of schooling could also be related to declines in the education gradient of HIV. To

In sum, the rebound of the HIV-Education gradient reflects both higher incidence of HIV infection (as a result of riskier sexual behavior) and higher survival rates for the stock of highly educated individuals (who can have better access to costly ART).

6.2 Empirical Evidence

Our previous theoretical interpretation suggests a parallel evolution between the HIV-Education gradient and education disparities in risky sexual behavior. We provide an empirical investigation of this phenomenon in this section.

The Risky Sex-Education Gradient The margins of risky sexual behavior that we study are: (i) the number of sex partners other than spouses (i.e., extramarital partners) during past 12 months, i.e., the extensive margin of sexual behavior,[26] and (ii) a dummy variable equal to 1 if the respondent used a condom during the last intercourse.We label the first Risky Sex-Education gradient as the Partners-Education gradient and the second as the Condom-Education gradient.

The results for the Partners-Education gradient are in panel A of Table 6. We report the results separately for women and for men, and we follow the same econometric specifications described for the HIV-Education gradient in Table 5. The more educated have significantly more sexual partners that the less educated. An additional year of schooling increases the chances of having an extramarital partners by 1.77% for women and by 2.85% for men (column (1) and (6), panel A, Table 5).[27] The non-stationary specification uncovers an interesting inverted U-shaped pattern of the Partners-Education gradient that represents our main finding in this section. The Partners-Education gradient first decreases (between aggregate stages 0 and 2) and then increases (between aggregate stages 2 and 4) for both women and men, and this dynamics are significant. The dynamics across stages of the epidemic show that the evolution of the Partners-Education gradient is remarkably consistent with the pattern of the HIV-Education gradient, as it is predicted by our theory in the previous section. To see this, Figure 8 shows separately for

determine whether this is the case requires an evaluation of the effects of HIV on schooling investments of children and teenagers (Fortson, 2011). These effects of HIV on education are likely to appear at later rather than at earlier stages of the epidemic for the simple reason that it takes years to reach adulthood—our study sample.

[26] For the extensive margin of risky sexual behavior we use the number of sex partners other than spouses (i.e., extramarital partners) in the past 12 months and note that for individuals who are single or do not cohabit, all sex partners are extramarital.

[27] Note that the females' Partners-Education gradient is smaller (about two thirds) than that of males but it shows a similar pattern over the stages of the epidemic. One potential caveat of this analysis is that women might under-report risky sexual behavior more than men, as it was pointed out by Smith (1992) and Gersovitz et al. (1998). However, more recently, Helleringer et al. (2009) find that men and women are equally likely to under-report risky sexual behavior when using sexual network data from Likoma Island, Malawi. In our analysis, as long as misreporting occurs systematically across all stages of the epidemic, the shape of the Partners-Education gradient for women over stages will not be biased.

women (panel A) and for men (panel B) the isomorphic representation of the HIV-Education gradient and the Partners-education gradient (as in Figure 5, the significance of the gradients is denoted by the color of the marker). The sizes of the gradients are different, being larger for the Partners-Education gradient, which implies a positive elasticity of less than one from risky sex to HIV. In contrast, the Condom-Education gradient is reported in panel B of Table 6. In the stationary specification (columns 1 and 4) we find that the more educated women and men use, on average, more condoms than the less educated ones. Interestingly, we do not find a significant pattern across stages of development in the Condom-Education gradient. After Stage 0, the Condom-Education gradient remains positive but relatively constant across stages of the epidemic for both women and men. Finally, knowledge about the transmission mechanisms of HIV might affect sexual behavior (Dinkelman et al., 2006; Dupas, 2011; Duflo et al., 2015). We document that more-educated individuals acquire more information about HIV transmission than less-educated individuals at earlier stages of the epidemic, but these educational differences in knowledge remain constant as the epidemic evolves. This way, while knowledge might affect the HIV-Education gradient at early stages of the epidemic, this effect should rapidly vanish after the first stages.[28]

To sum up, the evolution of the Partners-Education gradient is consistent with the evolution of the HIV-Education gradient, as predicted by the theory. The evolution of educational disparities in the number of extramarital partners help explain both the decline and rebound of the HIV-Education gradient across stages of the epidemic.

Antiretroviral Therapy (ART) The effects of ART on the HIV-Education gradient are potentially ambiguous. On the one hand, ART increases survival probabilities (Greenwood et al., 2013) and decreases the degree of infectiousness of the HIV+ population that takes ART (Apondi et al., 2011). These effects, respectively, an increase in γ^+ and a reduction in $\lambda(x)$, unambiguously increase sexual behavior by reducing the future marginal cost of today's risky sex (panel (b) and (c) in Appendix Figure A-3). On the other hand, if agents have to pay for this treatment they will reduce today's risky sex because of the higher future marginal cost.[29] As long as the monetary cost of ART does not offset the reductions in the marginal cost of sex through γ and λ, ART will increase risky sex (see panel (b) and (c) in Appendix Figure A-4).

[28] To this end, we consider two DHS questions regarding ways to avoid HIV infection that are directly related to the risky sex margins studied in the previous subsection. Specifically, respondents answer two questions: (i) *"Can you (the respondent) reduce the chances of getting HIV by having one sex partner who has no other partners?"* and (ii) *"Can you (the respondent) reduce the chances of getting HIV by always wearing a condom?"*. We estimate the education gradients in these knowledge variables. Our results are in appendix Table A-2.

[29] In our model, a monetary cost of ART would imply a change in the future marginal cost of having risky sex today, which becomes $\beta\lambda_x(\gamma^- u(y_1(s)) - \gamma^+ u(y_1(s) - cost))$.

Treatment is indeed costly. While prices of the most common first-line ART regiments have declined over time, they remain relatively high for a vast majority of the population, with median prices in low and medium income countries of USD115 per patient per year (ppy) in 2013. The prices of USD330 (ppy) in 2013 for the second-line treatment, and more than USD1,500 for the third-line treatment, (WHO, 2014). This is prohibitive for a vast majority of SSA households.[30] In this context, if the more educated have more access to ART , then ART might help explain the rebound in the most advanced stages. The findings in Table 7 point in this direction. We estimate the HIV-Education gradient by epidemiological stage with and without ART controls in panel A and B, respectively. Note that when we include ART, the estimate of HIV-Education gradient in the most advanced stages of the epidemic decreases with respect to the counterpart in panel A. This suggests that the provision of ART partially accounts for the rebound in the gradient. This result should be taken with a grain of salt though, since we cannot directly test the impact of the education gradient on HIV because the DHS does not provide information of ART at the individual level.

7 Conclusion

The mixed evidence in the literature investigating the relationship between education and the probability of being HIV-positive in SSA suggests that finding which type of individuals are at greater risk of HIV infection is not an easy task. We proposed a fresh look to this question that consists of explicitly introducing the stages of the HIV epidemic into the analysis. Using nationally representative data from 39 DHS surveys to exploit variation across stages of the HIV epidemic, we showed that the relationship between completed educational attainment and individual HIV status (i.e., the HIV-Education gradient) is dynamic, and significantly evolves with the epidemic. At early stages of the epidemic more-educated individuals are more likely to be infected; however, this relationship strongly decreases as the epidemic evolves, and eventually reaches a stage where education and the probability of being HIV-positive are no longer significantly correlated. Interestingly, in the most advanced stages of the epidemic, the education gradient of HIV returns to being high and positive. We showed theoretically and empirically that the educational disparities in risky sexual behavior (in terms of extramarital partners) closely resemble the U-shaped pattern of the education gradient in HIV.

[30] For example, income per capita in Malawi is on average USD250 in 2014, and the average income per capita in SSA is USD1638. We also expect more-educated individuals to have greater access to ART treatments for several obvious reasons—they (i) are more likely to live in the city (e.g., in Malawi, anyone who has a university degree is likely to live in the two largest cities, Lilongwe or Blantyre, where the ART drugs are available), (ii) have better transportation (do not have to walk several miles to refill prescriptions), or (iii) have access to someone in a hospital who can help them gain priority status when necessary to obtain ART .

In light of our findings, we call for frameworks of policy evaluation that incorporate the stylized dynamic relationship between education, HIV and risky sexual behavior that we document along the course of the HIV epidemic.

References

Aizer, A. and Stroud, L. (2010). Education, knowledge and the evolution of disparities in health. NBER Working Paper, 15840.

Anglewicz, P. and Kohler, H.-P. (2009). Overestimating hiv infection: The construction and accuracy of subjective probabilities of hiv infeciton in rural malawi. *Demography Research*, 20(6).

Apondi, R., Bunnell, R., Ekwaru, J. P., Moore, D., Bechange, S., Khana, K., King, R., Campbell, J., Tappero, J., and Mermin, J. (2011). Sexual behavior and hiv transmission risk of ugandan adults taking antiretroviral therapy: 3 year follow-up. *AIDS*, 10:1317–1327.

Asiimwe-Okiror, G., Opio, A. A., Musinguzi, J., Madraa, E., Tembo, G., and Caraël, M. (1997). Change in sexual behaviour and decline in hiv infection among young pregnant women in urban uganda. *AIDS*, 11(14):1757–63.

Bärnighausen, T., Hosegood, V., Timaeus, I. M., and Newell, M.-L. (2007). The socioeconomic determinants of hiv incidence: Evidence from a longitudinal, population-based study in rural south africa. *AIDS*, 21(7):S29–S38.

Beegle, K. and de Walque, D. (2009). Demographic and socioeconomic patterns of hiv/aids prevalence in africa. World Bank Policy Research Working Paper, 5076.

Behrman, J. and Kohler, H. (2012). Sexual Transmission of HIV. In Lomborg, B., editor, *Rethink HIV*, chapter 1. Cambridge University Press.

Bolt, J. and van Zanden, J. L. (2013). The First Update of the Maddison Project; Re-Estimating Growth Before 1820. *Maddison Project*, (4).

Bongaarts, J., Buettner, T., Heilig, G., and Pelletier, F. (2008). Has the hiv epidemic peaked? *Population and Development Review*, 34(2):199–224.

Boozer, M. and Philipson, T. (2000). The Impact of Public Testing for Human Immunodeficiency Virus. *The Journal of Human Resources*, 35(3):419–446.

Buera, F. J., Kaboski, J. P., and Rogerson, R. (2015). Skill biased structural change. Working Paper 21165, National Bureau of Economic Research.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3):414–427.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives*, 20(3):97–120.

Cutler, D. and Lleras-Muney, A. (2010). Understanding Differences in Health behaviors by Education. *The Journal of Health Economics*, 29(1):1–28.

Cutler, D. and Lleras-Muney, A. (2011). Education and Health. Working Paper.

Cutler, D. and Lleras-Muney, A. (2014). Education and Health: Insights from International Comparisons. In Culyer, A. J., editor, *Encyclopedia of Health Economics*, volume 1, pages 232–245. San Diego Elsevier.

De Magalhães, L. and Santaeulàlia-Llopis, R. (2015). The Consumption, Income, and Wealth of the Poorest: Cross-Sectional Facts of Rural and Urban Sub-Saharan Africa for Macroeconomists. World Bank, Policy Research Working Paper 7337.

de Walque, D. (2004). Education, information, and smoking decisions: evidence from smoking histories, 1940-2000. World Bank Policy Research Working Paper, 3362.

de Walque, D. (2007). How does the impact of an hiv/aids information campaign vary with educational attainment? evidence from rural uganda. *Journal of Development Economics*, 84(2):686–714.

de Walque, D. (2009). Does education affect hiv status? evidence from five african countries. *The World Bank Economic Review*, 23(2):209–233.

de Walque, D., Nakiyingi-Miiro, J. S., Busingye, J., and Whitworth, J. A. (2005). Changing association between schooling levels and hiv-1 infection over 11 years in rural populations cohort in south-west uganda. *Tropical Medicine and International Health*, 10(10):993–1001.

Delavande, A. and Kohler, H.-P. (2012). The Impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi. *Demography*, 49(3):1011–1036.

DePaula, A., Shapira, G., and Todd, P. E. (2014). How Beliefs About Hiv Status Affect Risky Behaviors: Evidence From Malawi. *Journal of Applied Econometrics*, 29(6):944–964.

Dinkelman, T., Levinsohn, J. A., and Majelantle, R. (2006). When knowledge is not enough: Hiv/aids information and risky behavior in botswana. NBER Working Paper, 12418.

Duflo, E., Dupas, P., and Kremer, M. (2015). Education, HIV, and Early Fertility: Experimental Evidence from Kenya. *American Economic Review*, 105(9):2757–97.

Dupas, P. (2011). Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics*, 3(1):1–36.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, 35:553–572.

Feldman, J. J., Makuc, D. M., Kleinman, J. C., and Cornoni-Huntley, J. (1989). National trends in educational differences in mortality. *Americal Journal of Epidemiology*, 10:214–23.

Fortson, J. G. (2008). The gradient in sub-saharan africa: Socioeconomic status and hiv/aids,. *Demography*, 45(2):303–322.

Fortson, J. G. (2011). Mortality risk and human capital investment: The impact of hiv/aids in sub-saharan africa. *Review of Economics and Statistics*, 93(1):1–15.

Galor, O. and Weil, D. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American Economic Review*, 90:806–828.

Gersovitz, M., Jacoby, H. G., Dedy, F. S., and Tap, A. G. (1998). The balance of self-reported heterosexual activity in kap surveys and the aids epidemic in africa. *Journal of the American Statistical Association*, 93:875–883.

Glynn, J. R., Caraël, M., Buvé, A., Anagonou, S., Zenkeng, L., Kahindo, M., and Musonada, R. (2002). Does increased general schooling protect against hiv infection? a study in four african cities. *Tropical Medicine and International Health*, 9:4–14.

Godlonton, S. and Thornton, R. L. (2013). Learning from Others' HIV Testing: Updating Beliefs and Responding to Risk. *American Economic Review*, 103(3):439–44.

Goldman, D. and Smith, J. P. (2005). Socioeconomic differences in the adoption of medical technology. *American Economic Review*, 95:234–37.

Gollin, D., Parente, S. L., and Rogerson, R. (2002). The role of agriculture in development. *American Economic Review*, 92(2):160–164.

Gollin, D., Parente, S. L., and Rogerson, R. (2007). The food problem and the evolution of international income levels. *Journal of Monetary Economics*, 54(4):1230–850.

Greenwood, J., Kircher, P., Santos, C., and Tertilt, M. (2013). An equilibrium model of the african hiv/aids epidemic. Working Paper 18953, National Bureau of Economic Research.

Greenwood, J., Seshadri, A., and Vandenbroucke, G. (2005). The baby boom and baby bust. *American Economic Review*, 95(1):183–207.

Gregson, S., Waddell, H., and Chandiwana, S. (2001). School education and hiv control in sub-saharan africa: From discord to armony. *Journal of International Development*, 13:467–485.

Hansen, G. D. and Prescott, E. C. (2002). Malthus to Solow. *American Economic Review*, 92(4):1205–1217.

Hargreaves, J. R., Bonell, C. P., Morison, L. A., Kim, J. C., Phetla, G., Porter, J. D., Watts, C., and Pronyk, P. M. (2007). Explaining continued high hiv prevalence in south africa: Socioeconomic factors, hiv incidence and sexual behavior change among a rural cohort, 2001-2004. *AIDS*, 21(7):S39–S48.

Hargreaves, J. R. and Glynn, J. R. (2002). Education attainment and hiv-1 infection in developing countries: A systematic review. *Tropical Medicine and International Health*, 7(6):489–498.

Helleringer, S., Kohler, H.-P., Chimbiri, A., Chatonda, P., and Mkandawire, J. (2009). The likoma network study: Context, data collection, and initial results. *Demography Research*, 21:427–468.

Herrendorf, B., Rogerson, R., and kos Valentinyi (2014). Chapter 6 - growth and structural transformation. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, pages 855 – 941. Elsevier.

Kates, J., Wexler, A., Lief, E., Avila, C., and Gobet, B. (2011). Financing the response to aids in low- and middle- income countries: International assistance from donor governments in 2010. Joint United Nations Programme on HIV/AIDS and The Henry J. Kaiser Foundation.

Lee, R. (2003). The Demographic Transition: Three Centuries of Fundamental Change. *Journal of Economic Perspectives*, 17(4):167–190.

Link, B. G., Northridge, M. E., Phelan, J. C., and Ganz, M. L. (1998). Social epidemiology and the fundamental cause concept: On the structuring of effective cancer screens by socioeconomic status. *Milbank Quarterly*, 76(3):375–402.

Link, B. G. and Phelan, J. C. (1995). Social conditions as the fundamental causes of disease. *Journal of Health and Social Behavior*, 35:80–94. Extra Issue: Forty Years of Medical Sociology: The State of the Art and Directions for the Future.

Lleras-Muney, A. (2004). The relationship between education and adult mortality in the united states. *Review of Economic Studies*, 72(1):189–221.

Lopman, B., Lewis, J., Nyamukapa, C., Mushati, P., Chandiwana, S., and Gregson, S. (2007). Hiv incidence and poverty in manicaland, zimbabwe: Is hiv becoming the disease of the poor? *AIDS*, 21(7):S57–S66.

Lucas, R. E. (2004). The Industrial Revolution: past and future. *Annual Report*, (May):5–20.

Mishra, V., Bignami-Van-Assche, S., Greener, R., Vaessen, M., Hong, R., Ghys, P. D., Boerma, J. T., Van-Assche, A., Shane-Khan, and Rutstein, S. (2007). Hiv infection does not disproportionately affect the poorer in sub-saharan africa. *AIDS*, 21(7):S17–S29.

Nyqvist, M. B., Corno, L., De Walque, D. B. C. M., and Svensson, J. (2015). Using Lotteries to Incentivize Safer Sexual Behavior: Evidence from a Randomized Controlled Trial on HIV Prevention. Policy Research Working Paper Series 7215, The World Bank.

Preston, S. H. and Elo, I. T. (1995). Are educational differentials in adult mortality increasing in the u.s.? *Journal of Aging Health*, 7(4):476–496.

Smith, T. W. (1992). A methodological analysis of the sexual behavior questions on the general social surveys. *Journal of Official Statistics*, 8:309–325.

Stoneburner, R. L. and Low-Beer, D. (2004). Population-level hiv declines and behavioral risk avoidance in uganda. *Science*, 304:714–8.

Strauss, J. and Thomas, D. (2007). Health over the life course. In Schultz, T. P. and Strauss, J., editors, *Handbook of Development Economics*, volume 4, chapter 54, pages 3375–3474. Amsterdam, North-Holland.

Thornton, R. L. (2008). The demand for, and impact of, learning hiv status. *American Economic Review*, 98(5):1829–63.

Thornton, R. L. (2012). HIV Testing, Subjective Beliefs and Economic Behavior. *Journal of Development Economics*, 99(2):300–313.

UNAIDS (1997). A measure of success in uganda: the value of monitoring both hiv prevalence and sexual behavior. Geneva: UNAIDS.

UNAIDS (2015). Report on the global aids epidemic. Joint United Nations Programme on HIV/AIDS, New York.

WHO (2014). Access to Antiretroviral Drugs.

Wojcicki, J. M. (2005). Socioeconomic status as a risk factor for hiv infection in women in east, central and southern africa: A systematic review. *Journal of Biosocial Science*, 37(1):1–36.

Table 1: The DHS Sample Characteristics (across Countries)

(A) Women	Mean	Median	Min.	Max.	Gini
HIV Prevalence (%)	6.1	4.1	0.5	31.2	0.54
Years of Schooling	3.2	2.8	0.6	4.8	0.26
Age	28.2	28.2	27.7	29.4	0.01
Urban (%)	32.7	33.6	10.6	88.3	0.25
Extramarital Partners, n	0.15	0.17	0.01	0.56	0.37
$n = 0$ (%)	87.4	84.9	57.9	99.2	0.07
$n = 1$ (%)	11.5	14.3	0.8	35.9	0.36
$n = 2$ (%)	1.0	1.2	0.0	6.3	0.47
$n \geq 3$ (%)	0.1	0.1	0.0	0.0	0.66
Frequency of Condom Use (%)	9.5	8.7	0.5	37.7	0.42
Not Sexually Active (%)	15.6	12.8	4.3	30.7	0.28

(B) Men	Mean	Median	Min.	Max.	Gini
HIV Prevalence (%)	4.1	2.2	0.4	19.7	0.56
Years of Schooling	3.8	3.2	1.3	5.3	0.18
Age	28.2	28.5	25.9	30.3	0.02
Urban (%)	33.7	35.6	15.4	87.6	0.24
Extramarital Partners, n	0.41	0.48	0.08	1.17	0.30
$n = 0$ (%)	73.7	65.7	36.8	93.5	0.14
$n = 1$ (%)	19.7	28.0	5.1	43.8	0.22
$n = 2$ (%)	4.5	5.2	0.8	14.4	0.36
$n \geq 3$ (%)	0.7	0.7	0.1	2.8	0.42
Frequency of Condom Use (%)	20.5	21.4	4.2	49.3	0.31
Not Sexually Active (%)	21.7	17.9	7.0	35.9	0.24

Notes: The computation of these statistics is performed by using individual HIV weights provided by the DHS to compute the mean. Then, country-specific population weights (i.e., the population size of each country provided by World Population Prospects) are used to compute the statistics across countries. The number of extramarital partners refers to the last 12 months. The frequency of condom use refers to the last sexual intercourse. Our sample is based on 39 DHS with 25 countries and a total of 227,935 women and 174,852 men.

Table 2: The Evolution of the HIV Epidemic across Sub-Saharan Countries: The DHS Sample

	DHS Obs.		Peak		DHS/Peak	
Country	t_i	HIV_i	$t_{i,*}$	$HIV_{i,*}$	$t_i - t_{i,*}$	$\frac{HIV_i}{HIV_{i,*}}$
Burkina Faso	2003	1.76	1993	3.96	-10	0.44
Burkina Faso	2010	1.10	1993	3.96	-17	0.27
Burundi	2010	1.39	1996	4.47	-14	0.31
Cameroon	2004	4.98	2002	5.06	-2	0.98
Cameroon	2011	4.47	2002	5.06	-9	0.88
Congo Brazaville	2009	2.74	1995	4.80	-14	0.57
Cote d'Ivoire	2005	4.79	1999	6.18	-6	0.77
Cote d'Ivoire	2011-12	3.41	1999	6.18	-12	0.55
D.R. Congo	2007	1.30	2001	1.43	-6	0.91
Ethiopia	2005	2.28	1999	3.23	-6	0.70
Ethiopia	2011	1.32	1999	3.23	-12	0.41
Gabon	2012	4.16	2003	5.53	-9	0.75
Ghana	2003	2.00	2000	2.15	-3	0.93
Guinea	2005	1.43	2010	1.54	5	0.93
Guinea	2012	1.52	2010	1.54	-2	0.98
Kenya	2003	7.08	1996	9.16	-7	0.77
Kenya	2008-09	5.84	1996	9.16	-12	0.64
Lesotho	2004-05	23.33	2010	23.91	6	0.98
Lesotho	2009-10	23.52	2010	23.91	1	0.98
Liberia	2006-07	1.34	2002	1.84	-5	0.73
Malawi	2004	13.41	1998	14.99	-6	0.89
Malawi	2010	11.21	1998	14.99	-12	0.75
Mali	2006	1.34	1999	1.71	-7	0.78
Mozambique	2009	9.98	2008	10.01	-1	1.00
Niger	2006	1.00	2002	1.20	-4	0.83
Niger	2012	0.60	2002	1.20	-10	0.50
Rwanda	2005	3.09	1994	5.93	-11	0.52
Rwanda	2010	2.90	1994	5.93	-16	0.49
Senegal	2005	0.90	2004	0.90	-1	1.00
Senegal	2010-11	0.70	2004	0.90	-6	0.78
Sierra Leone	2008	1.66	2008	1.66	0	1.00
Swaziland	2006-07	23.36	2010	24.46	4	0.95
Tanzania	2003-04	6.04	1996	7.36	-7	0.82
Tanzania	2007-08	5.30	1996	7.36	-11	0.72
Tanzania	2011-12	4.81	1996	7.36	-15	0.65
Uganda	2011	6.55	1991	12.62	-20	0.52
Zambia	2007	12.87	1998	14.77	-9	0.87
Zimbabwe	2005-06	17.46	1997	25.90	-8	0.67
Zimbabwe	2010-11	14.35	1997	25.90	-13	0.55

Notes: t_i is the calendar year of DHS data collection for country i; HIV_i is the prevalence rate for country i the year of DHS data collection; $t_{i,*}$ is the year country i reaches its HIV prevalence peak; $HIV_{i,*}$ is the peak prevalence rate for country i. Sources: United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects: The 2015 Revision, Medium-Variant Estimation and Projection.

Table 3: The HIV-Education Gradient

(A) *HIV Status*	(1)	(2)	(3)	(4)	(5)
Education	0.0043***	0.0112***	0.0098***	0.0040***	0.0037***
	(0.0006)	(0.0007)	(0.0010)	(0.0003)	(0.0003)
Education * Stage1		-0.0059***	-0.0046***	-0.0010***	-0.0008**
		(0.0008)	(0.0010)	(0.0003)	(0.0003)
Education * Stage2		-0.0116***	-0.0103***	-0.0027***	-0.0024***
		(0.0007)	(0.0011)	(0.0002)	(0.0003)
Education * Stage3		-0.0093***	-0.0076***	-0.0020*	-0.0018
		(0.0015)	(0.0011)	(0.0011)	(0.0011)
Education * Stage4		-0.0064***	-0.0051***	-0.0015***	-0.0012***
		(0.0008)	(0.0012)	(0.0003)	(0.0003)
Male	-0.0224***	-0.0229***	-0.0228***	-0.0223***	-0.0224***
	(0.0027)	(0.0027)	(0.0021)	(0.0020)	(0.0019)
Age	0.0025***	0.0025***	0.0025***	0.0025***	0.0025***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0002)
Urban Area	0.0212***	0.0197***	0.0227***	0.0280***	0.0285***
	(0.0044)	(0.0045)	(0.0040)	(0.0027)	(0.0027)
Stage 1	-0.0023	0.0124***	0.0111***	-0.0055***	0.0088***
	(0.0048)	(0.0036)	(0.0041)	(0.0008)	(0.0019)
Stage 2	0.0103	0.0498***	0.0598***	-0.0012	0.0200***
	(0.0078)	(0.0089)	(0.0100)	(0.0025)	(0.0032)
Stage 3	-0.0094	0.0197	0.0314***	-0.0102**	0.0110**
	(0.0128)	(0.0122)	(0.0096)	(0.0052)	(0.0043)
Stage 4	-0.0032	0.0131***	0.0394***	-0.0147***	0.0032
	(0.0042)	(0.0030)	(0.0072)	(0.0019)	(0.0029)
Agricultural Share	-0.0029***	-0.0029***	-0.0031***	0.0021***	-0.0008***
	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.0003)
Output per Capita	-0.0000***	-0.0000***	-0.0000***	0.0000**	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.0806***	0.0615***	0.0307***	-0.1363***	-0.1832***
	(0.0089)	(0.0072)	(0.0089)	(0.0344)	(0.0175)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	402,766	402,766	402,766	402,766	402,766

(B) *HIV Status*	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Education	0.0038***	0.0022***	0.0014***	0.0020*	0.0025***
	(0.000)	(0.000)	(0.000)	(0.0011)	(0.000)
Year-Country Dum.	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes
Sample Size	66,322	119,700	48,615	50,535	118,425

Notes: All specifications use the "Full Sample" described in Section 3 and the same set of controls. In Panel (A), Column (1) reports the results for the stationary specification, and columns (2) to (5) report the results for the non-stationary specification. We add year dummies in column (3), country dummies in column (4) and year-country dummies in column (5). Panel (B) reports the estimates of the HIV-Education gradient for each stage separately. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Additional Inference

(A) *HIV-Education Gradient*	(1)	(2)	(3)	(4)	(5)
γ_0	0.0043***	0.0112***	0.0098***	0.0040***	0.0037***
	(0.0006)	(0.0007)	(0.000)	(0.0003)	(0.0003)
$\gamma_0 + \gamma_1$		0.0053***	0.0052***	0.0029***	0.0029***
		(0.0005)	(0.000)	(0.0003)	(0.0003)
$\gamma_0 + \gamma_2$		-0.0004	-0.0005	0.0013***	0.0013***
		(0.0003)	(0.228)	(0.0001)	(0.0001)
$\gamma_0 + \gamma_3$		0.0019	0.0022***	0.0020*	0.0019*
		(0.0014)	(0.005)	(0.0011)	(0.0011)
$\gamma_0 + \gamma_4$		0.0048***	0.0047***	0.0025***	0.0025***
		(0.0005)	(0.000)	(0.0002)	(0.0001)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes

(B) *Rebound*	(1)	(2)	(3)	(4)	(5)
$\gamma_4 - \gamma_2$		0.0052***	0.0051***	0.0011***	0.0012***
		(0.0006)	(0.0005)	(0.0001)	(0.0001)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes

Notes: The underlying econometric models are as specified in the columns of Table 3. Column (1) reports the tests results for the stationary specification. Columns (2) to (5) report the tests results for the non-stationary specification. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: The HIV-Education Gradient: Women and Men Separately

	(A) Women					(B) Men				
HIV Status	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0048***	0.0148***	0.0129***	0.0051***	0.0048***	0.0033***	0.0076***	0.0068***	0.0028***	0.0026***
	(0.0009)	(0.0012)	(0.0017)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0005)	(0.0003)	(0.0004)
Education * Stage1		-0.0087***	-0.0070***	-0.0017***	-0.0014***		-0.0037***	-0.0032***	-0.0012***	-0.0010**
		(0.0013)	(0.0016)	(0.0004)	(0.0004)		(0.0007)	(0.0007)	(0.0004)	(0.0005)
Education * Stage2		-0.0146***	-0.0129***	-0.0033***	-0.0030***		-0.0089***	-0.0081***	-0.0024***	-0.0022***
		(0.0013)	(0.0018)	(0.0004)	(0.0004)		(0.0005)	(0.0006)	(0.0004)	(0.0004)
Education * Stage3		-0.0133***	-0.0115***	-0.0040**	-0.0038**		-0.0051***	-0.0037***	0.0003	0.0005
		(0.0025)	(0.0022)	(0.0018)	(0.0017)		(0.0008)	(0.0007)	(0.0008)	(0.0008)
Education * Stage4		-0.0094***	-0.0078***	-0.0028***	-0.0024**		-0.0041***	-0.0034***	-0.0010***	-0.0007**
		(0.0016)	(0.0021)	(0.0004)	(0.0004)		(0.0004)	(0.0006)	(0.0003)	(0.0003)
Age	0.0024***	0.0025***	0.0025***	0.0024***	0.0024***	0.0026***	0.0026***	0.0026***	0.0026***	0.0026***
	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0002)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Urban Area	0.0282***	0.0262***	0.0300***	0.0362***	0.0369***	0.0129***	0.0121***	0.0142***	0.0186***	0.0188***
	(0.0059)	(0.0063)	(0.0055)	(0.0039)	(0.0038)	(0.0030)	(0.0030)	(0.0026)	(0.0019)	(0.0019)
Stage 1	-0.0053	0.0133***	0.0137**	-0.0051***	0.0112***	0.0016	0.0128***	0.0108***	-0.0035***	0.0081**
	(0.0060)	(0.0048)	(0.0055)	(0.0010)	(0.0022)	(0.0037)	(0.0024)	(0.0029)	(0.0012)	(0.0032)
Stage 2	0.0107	0.0536***	0.0655***	-0.0016	0.0243***	0.0103*	0.0454***	0.0542***	-0.0007	0.0145***
	(0.0099)	(0.0111)	(0.0133)	(0.0032)	(0.0048)	(0.0057)	(0.0060)	(0.0062)	(0.0032)	(0.0034)
Stage 3	-0.0102	0.0271**	0.0411***	-0.0068	0.0182**	-0.0079	0.0092	0.0191***	-0.0159***	0.0006
	(0.0164)	(0.0137)	(0.0125)	(0.0063)	(0.0076)	(0.0092)	(0.0094)	(0.0063)	(0.0038)	(0.0034)
Stage 4	-0.0051	0.0155***	0.0466***	-0.0157***	0.0071*	-0.0009	0.0117***	0.0332***	-0.0112***	0.0012
	(0.0055)	(0.0041)	(0.0099)	(0.0023)	(0.0042)	(0.0028)	(0.0019)	(0.0049)	(0.0020)	(0.0053)
Agricultural Share	-0.0032***	-0.0032***	-0.0034***	0.0018***	-0.0014***	-0.0025***	-0.0025***	-0.0027***	0.0027***	0.0001
	(0.0003)	(0.0002)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.0001)	(0.0002)	(0.0004)	(0.0003)
Output per Capita	-0.0000***	-0.0000***	-0.0000***	0.0001***	0.0001***	-0.0000***	-0.0000***	-0.0000***	0.0000	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.0922***	0.0681***	0.0357***	-0.1902***	-0.2129**	0.0459***	0.0321***	0.0030	-0.0899**	-0.1642***
	(0.0112)	(0.0087)	(0.0111)	(0.0435)	(0.0879)	(0.0088)	(0.0076)	(0.0093)	(0.0373)	(0.0193)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	227935	227935	227935	227935	227935	174831	174831	174831	174831	174831

Notes: All specifications use the "Full Sample" described in Section 3. Column (1) and (6) report the results for the stationary specification respectively for women and men. Columns (2) and (7) report the results for the non-stationary specification respectively for women and men. We add year dummies in columns (3) and (8), country dummies in columns (4) and (9) and year-country dummies in column (5) and (10) respectively for women and men. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: The Risky Sex-Education Gradient: Women and Men Separately

(A) Number of Extramarital Partners

	Women					Men				
Risky Sex	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0177**	0.0372***	0.0348***	0.0294***	0.0296***	0.0285***	0.0304***	0.0269***	0.0290***	0.0295***
	(0.029)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
Education * Stage1		-0.0163**	-0.0222***	-0.0204***	-0.0199***		0.0046	-0.0071	-0.0167**	-0.0167**
		(0.025)	(0.000)	(0.006)	(0.008)		(0.742)	(0.515)	(0.050)	(0.029)
Education * Stage2		-0.0254***	-0.0267***	-0.0243***	-0.0249***		-0.0200	-0.0238***	-0.0290***	-0.0294***
		(0.006)	(0.000)	(0.003)	(0.003)		(0.125)	(0.007)	(0.000)	(0.000)
Education * Stage3		-0.0308***	-0.0288***	-0.0201**	-0.0213**		-0.0069	-0.0047	0.0010	0.0003
		(0.002)	(0.000)	(0.019)	(0.018)		(0.550)	(0.588)	(0.932)	(0.980)
Education * Stage4		-0.0152	-0.0134	-0.0134*	-0.0138*		-0.0010	0.0019	0.0021	0.0018
		(0.107)	(0.185)	(0.081)	(0.087)		(0.928)	(0.845)	(0.840)	(0.866)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	227,935	227,935	227,9358	227,935	227,935	174,831	174,831	174,831	174,831	174,831

(B) Condom Use in Last Intercourse

	Women					Men				
Risky Sex	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0079***	0.0192***	0.0193***	0.0125***	0.0126***	0.0066***	0.0129***	0.0141***	0.0162***	0.0166***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education * Stage1		-0.0102***	-0.0114***	-0.0054**	-0.0055***		-0.0038	-0.0054	-0.0064*	-0.0066*
		(0.001)	(0.142)	(0.011)	(0.002)		(0.0406)	(0.122)	(0.084)	(0.065)
Education * Stage2		-0.0135***	-0.0132***	-0.0075***	-0.0075***		-0.0109***	-0.0111**	-0.0149***	-0.0156***
		(0.000)	(0.000)	(0.000)	(0.000)		(0.010)	(0.028)	(0.000)	(0.000)
Education * Stage3		-0.0138	-0.0133	-0.0068	-0.0068		-0.0101	-0.0108	-0.0131	-0.0136
		(0.164)	(0.000)	(0.482)	(0.501)		(0.611)	(0.640)	(0.592)	(0.585)
Education * Stage4		-0.0134**	-0.0133***	-0.0073**	-0.0080**		-0.0078**	-0.0088*	-0.0066***	-0.0074***
		(0.000)	(0.000)	(0.024)	(0.007)		(0.017)	(0.062)	(0.004)	(0.001)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	163,883	163,883	163,883	163,883	163,883	120,840	120,840	120,840	120,840	120,840

Notes: In panel (A) we report the marginal effects of the associated Tobit model where the endogenous variable is the number of extramarital partners in the past 12 months. In panel (B) we report the coefficients of a linear model where the endogenous variable is binary and refers to use of condom in last sexual intercourse. In both panels we include the same set of controls and fixed effects as in our benchmark specifications in Table 5. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis.* significant at 10%; ** significant at 5%; *** significant at 1%.

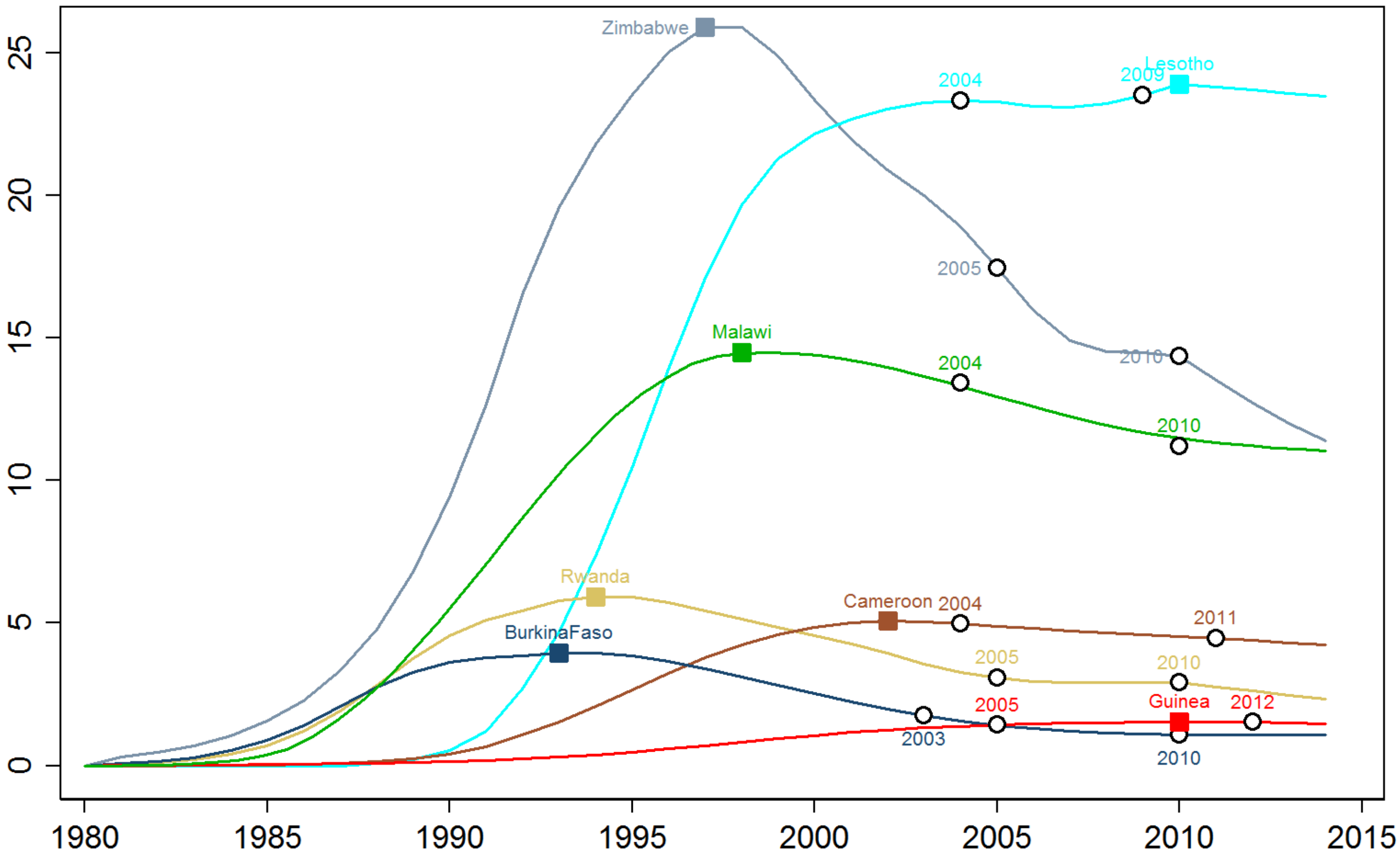
Table 7: The HIV-Education Gradient: Specification by Stage

(A) *HIV Status*	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Education	0.0044***	0.0022***	0.0014***	0.0020*	0.0025***
	(0.000)	(0.000)	(0.000)	(0.081)	(0.000)
Year-Country Dum.	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes
Sample Size	58,560	112,024	48,615	50,535	118,425

(B) *HIV Status with ART*	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Education	0.0043***	0.0029***	0.0016***	0.0011	0.0015***
	(0.000)	(0.0003)	(0.0002)	(0.0021)	(0.0002)
ART Coverage	0.0003*	-0.0025***	0.0015***	-0.0012***	-0.0003***
	(0.0002)	(0.0005)	(0.0000)	(0.0001)	(0.0000)
Education * ART Coverage	-0.0000	-0.0000	-0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Year-Country Dum.	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes	Yes-Yes
Sample Size	58,560	112,024	48,615	50,535	118,425

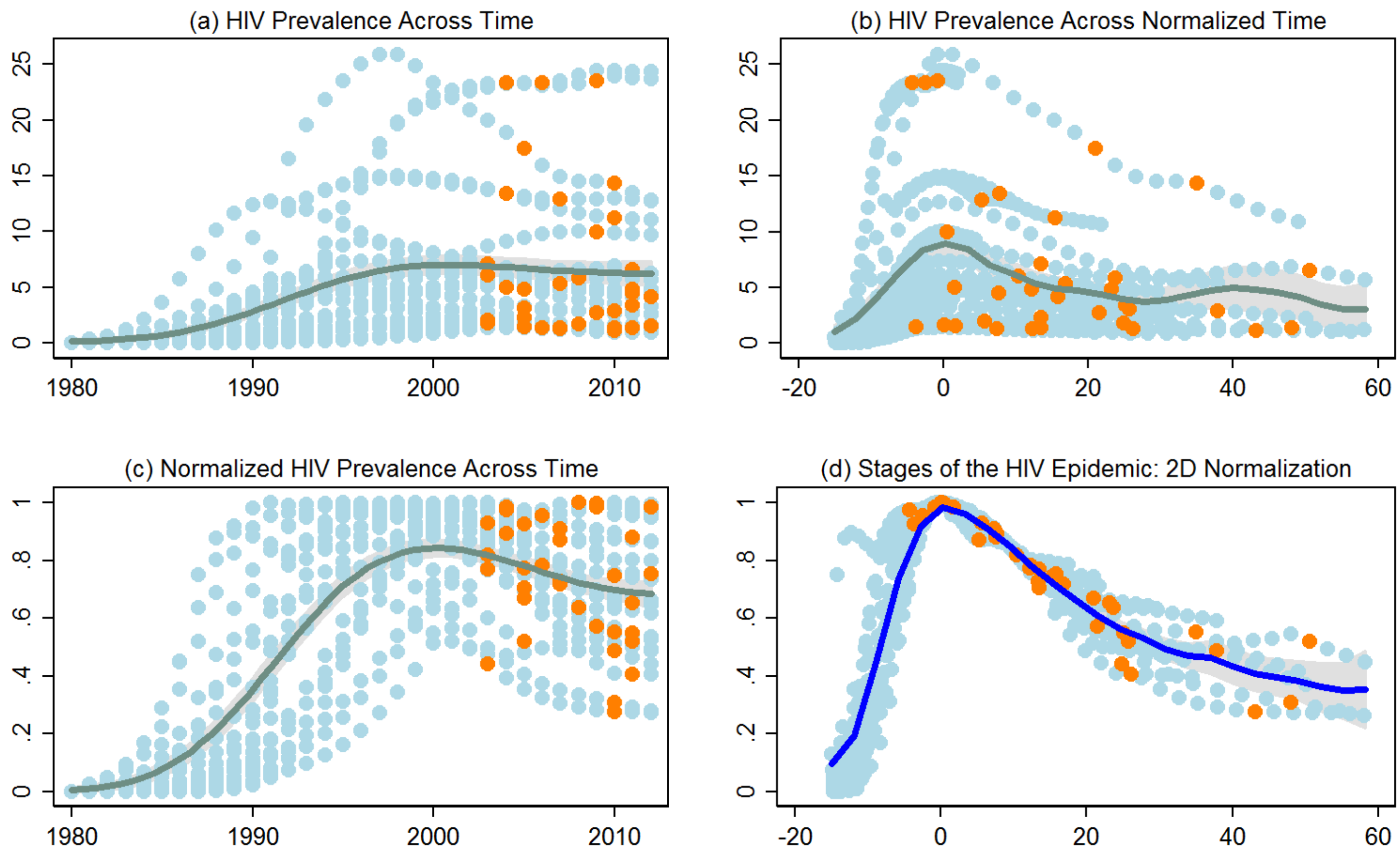
Notes: We apply our stationary specification of the HIV-Education gradient separately for each stage of the epidemic. We include year and country fixed effects in all columns. In both panels we include the same set of controls as in our benchmark specifications in Table 3. We exclude Senegal (in Stage 0) and Niger (in Stage 1) since WPP does not provide information about ART coverage for these countries. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Figure 1: Challenges for the Definition of Stages of the HIV: Epidemic. The Evolution of the HIV Epidemic for a DHS Subsample.



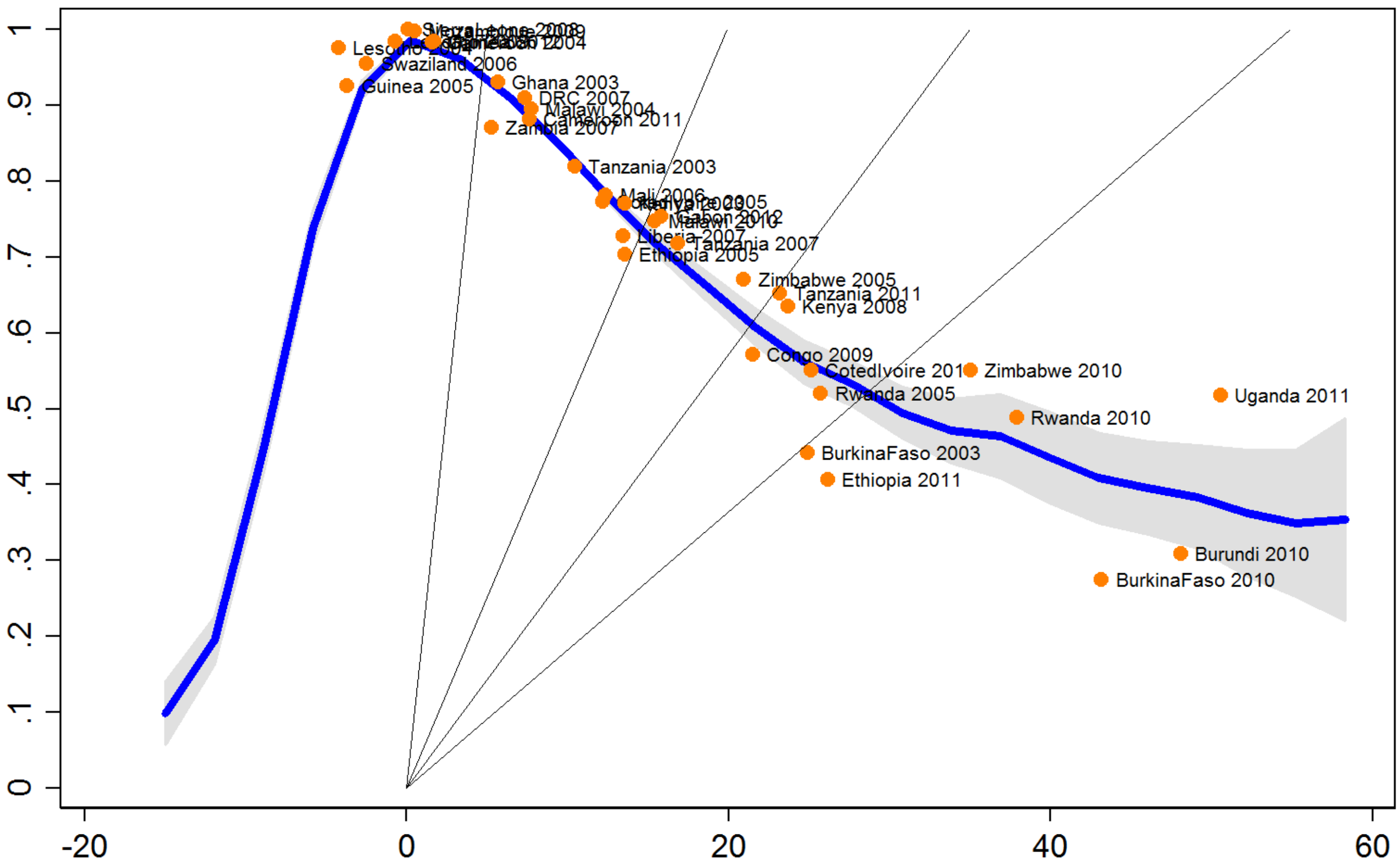
Source: United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects: The 2015 Revision, Medium-Variant Estimation. *Notes*: The solid square on each HIV time path displays the HIV prevalence at the peak year, and the open black circle on each HIV time path displays the HIV prevalence at the year that the DHS data were collected.

Figure 2: Defintion of Stages of the HIV Epidemic



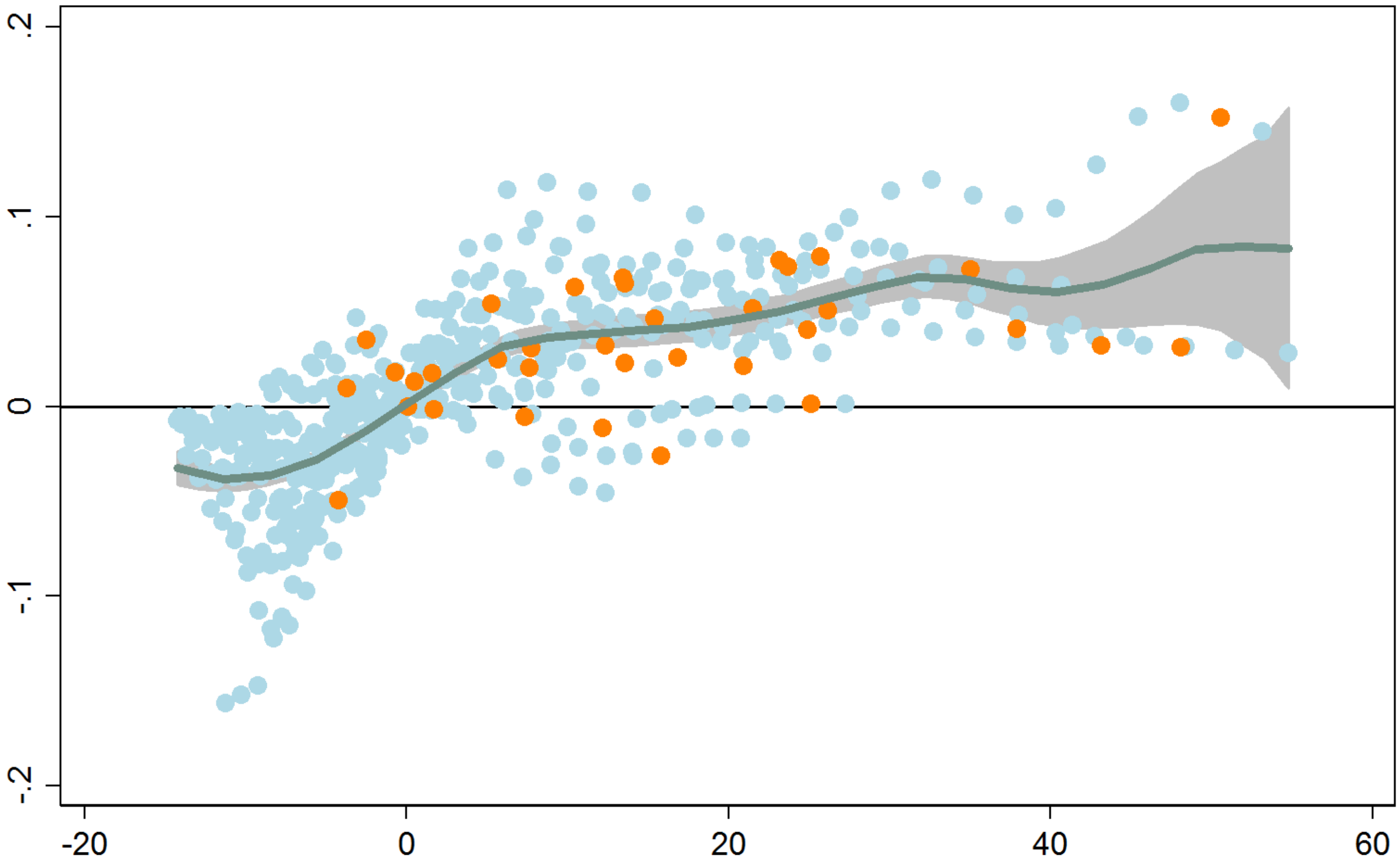
Source: Outcome of our 2D-normalization algorithm (Subsection 4.2) implemented using WPP, 2015, data for SSA. The vertical axis in the top panels is HIV prevalence. The vertical axis in the bottom panels is normalized HIV prevalence. The horizontal axis in the left panels is time. The horizontal axis in the right panels is normalized time. This way, the 2D-normalization is operative in panel (d) (Subsection 4.2). In each panel, the orange markers in the scatterplots represent a DHS dataset. The plotted trends are locally weighted polynomials with 95% confidence intervals.

Figure 3: Stages of the HIV Epidemic: DHS Sample



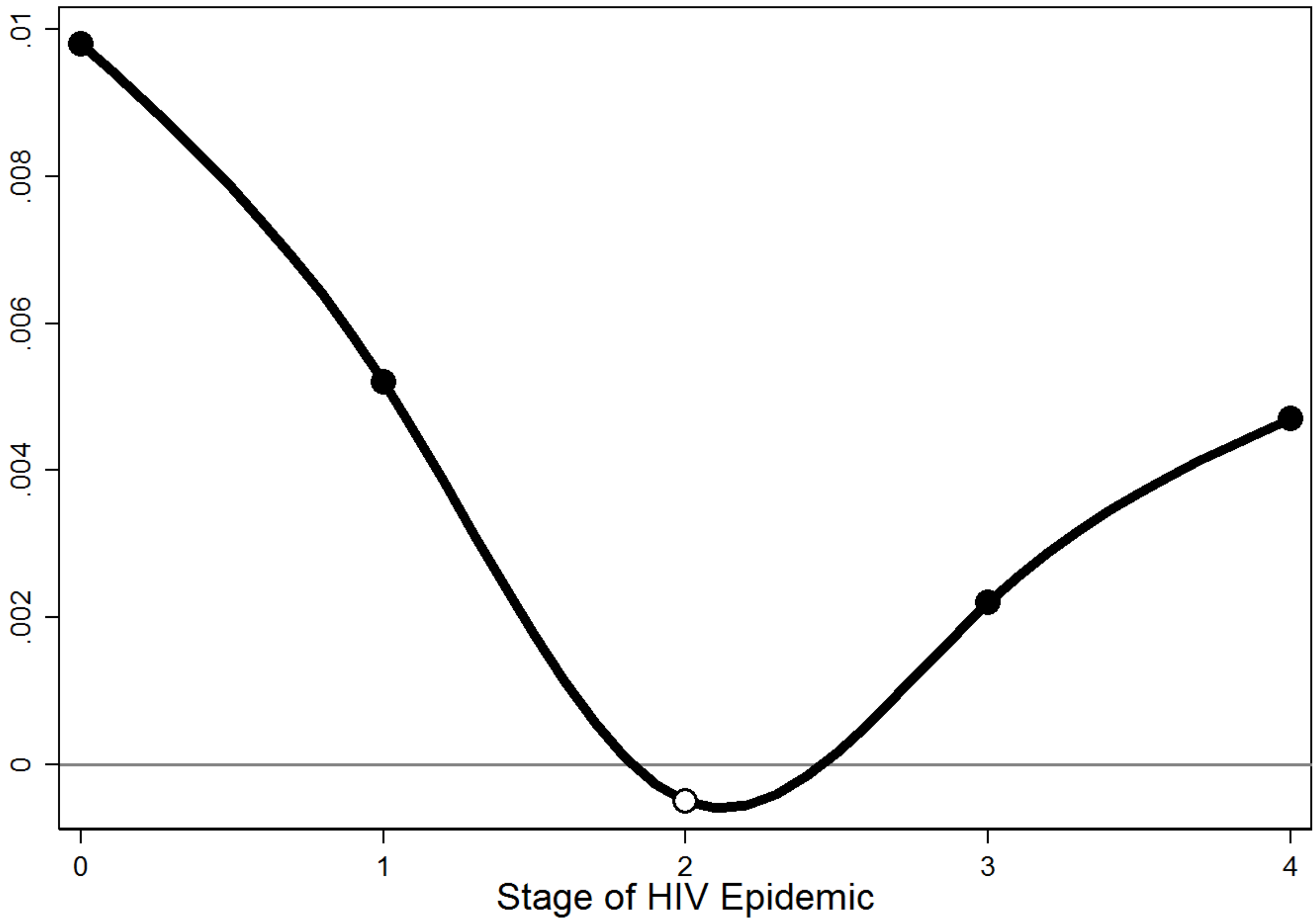
Source: The plot shows the location of SSA Countries (DHS Sample) on the 2D-normalized space at the time of DHS data collection. The vertical axis is the normalized HIV prevalence, and the horizontal axis is the normalized time (Subsection 4.2). The orange markers in the scatterplots represent a DHS dataset. The plotted trends are locally weighted polynomials with 95% confidence intervals.

Figure 4: HIV Gender Gap (Women - Men) Across Stages of the Epidemic, DHS and SSA Sample



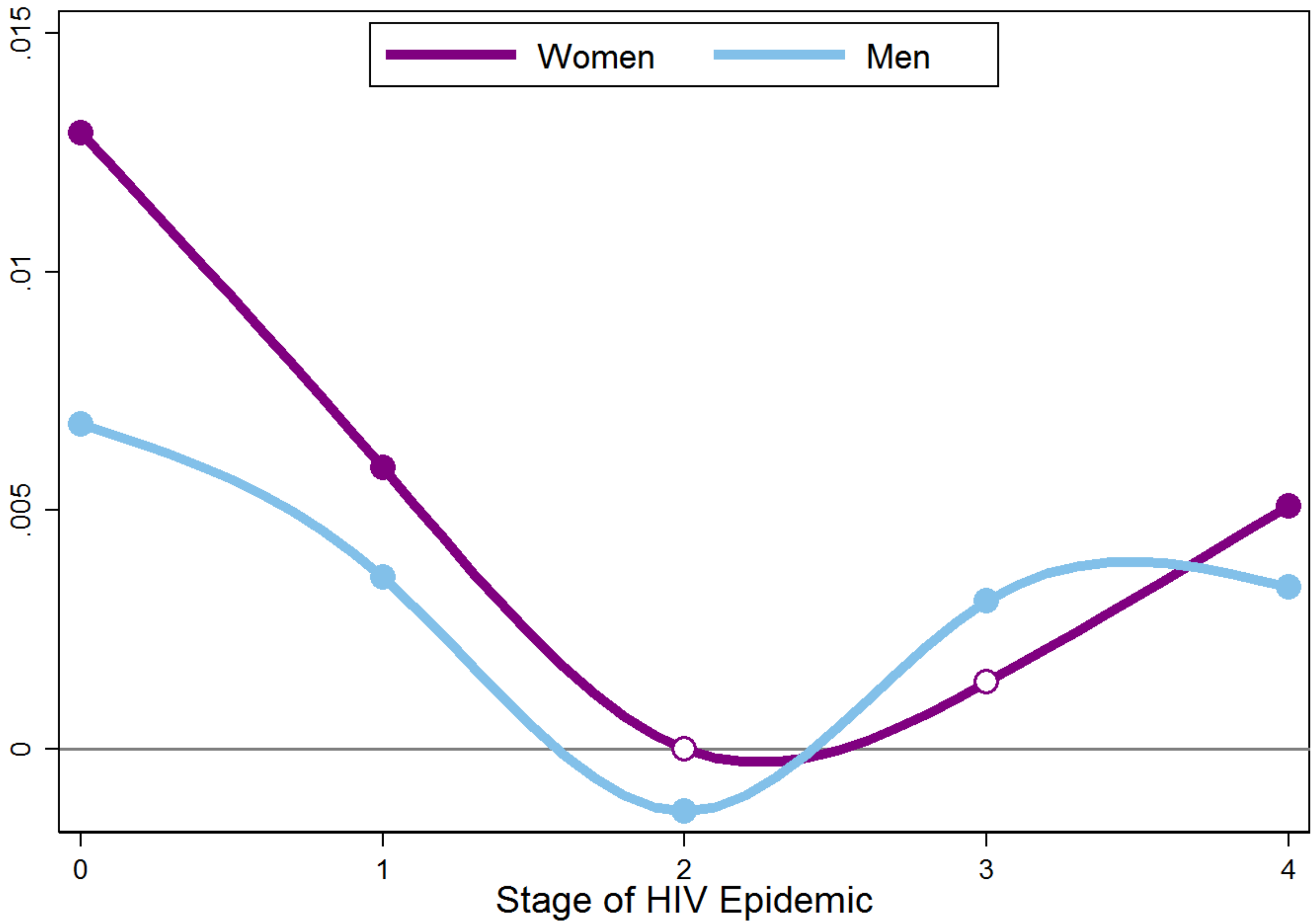
Source: The plot shows the HIV gender gap (i.e., HIV prevalence of women minus HIV prevalence of men) across the normalized time using WPP, 2015, data for SSA. The vertical axis is the normalized HIV prevalence, and the horizontal axis is the normalized time (Subsection 4.2). The orange markers in the scatterplots represent a DHS dataset. The plotted trends are locally weighted polynomials with 95% confidence intervals.

Figure 5: The HIV-Education Gradient Across Stages of the Epidemic



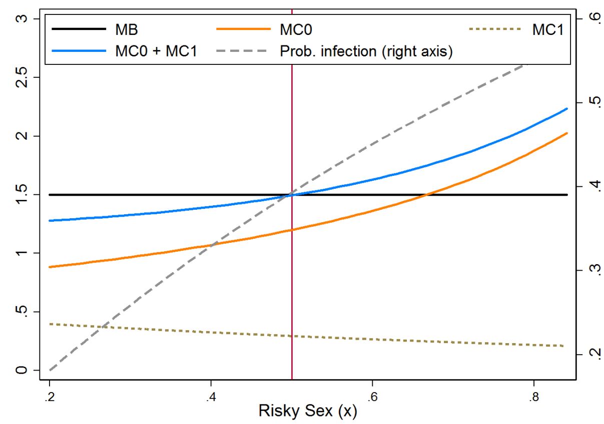
Notes: This graph plots the benchmark estimates of the HIV-Education gradient using the full sample (with year controls). For each stage j we plot $\left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j\right)$. We construct this estimates from column 3 of Table 3 (also reported in column 3, panel A, Table 4). Significance at 10%, 5%, and 1% is represented by, respectively, markers with open circles, markers with medium transparency fill, and markers with solid fill. We use a cubic spline for interpolation across stages.

Figure 6: The HIV-Education Gradient: Women and Men Separately

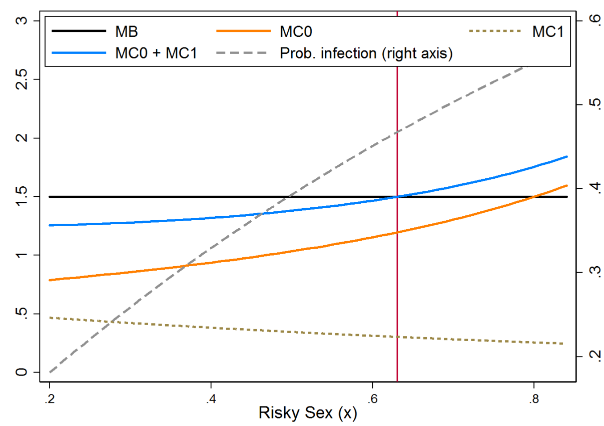


Notes: This graph plots the benchmark estimates of the HIV-Education gradient for each stage j for women and men separately (with year controls). For each stage j we plot $\left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j\right)$. For women (men) we construct this gradient using the estimates in column 3 (8) in Table 5. Significance at 10%, 5%, and 11% is represented by, respectively, markers with open circles, markers with medium transparency fill, and markers with solid fill. We use a cubic spline for interpolation across stages.

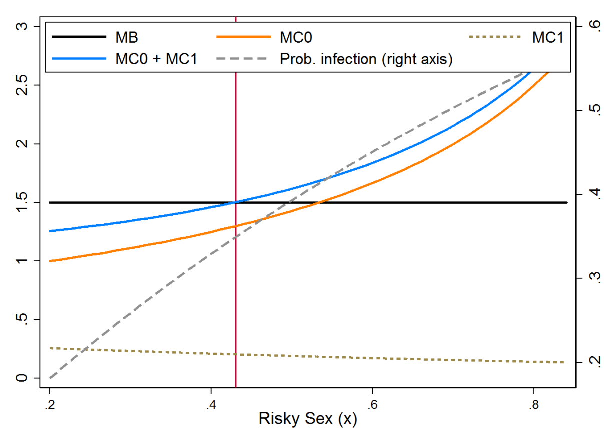
Figure 7: Theoretical Interpretation: The Risky Sex-Education Gradient



(a) Benchmark



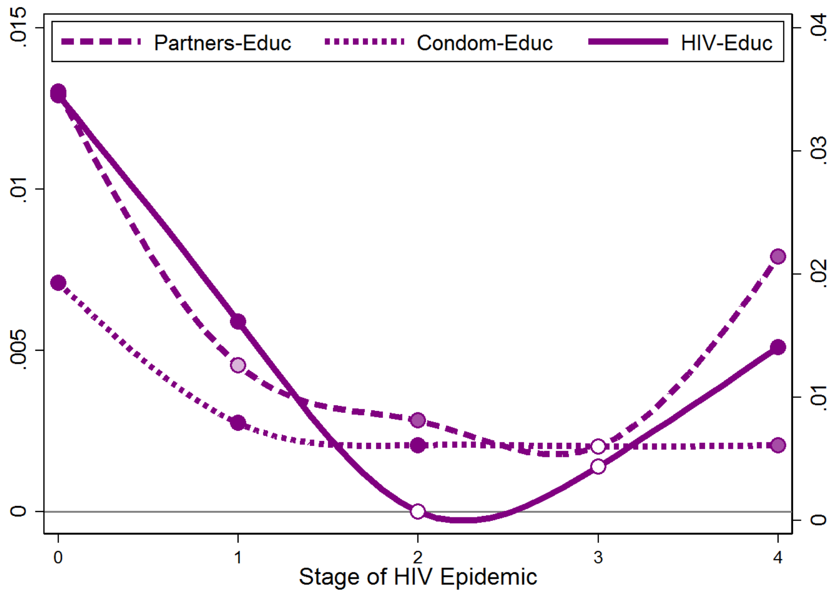
(b) High Education



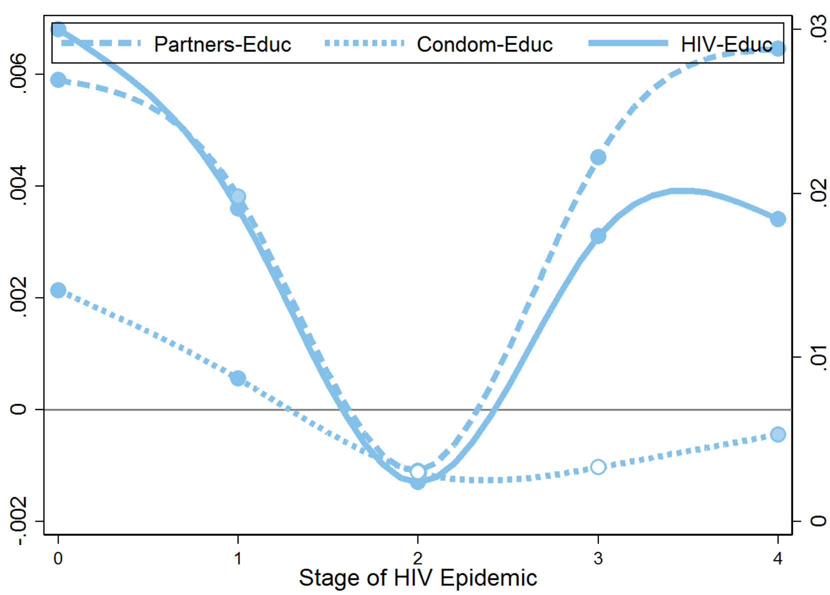
(c) Low Education

Notes: The horizontal axis denotes risky sex, x. All variables are plotted on the left vertical axis, except for the probability of infection $\lambda(x)$ that is plotted on the right axis. We denote the marginal benefit as MB, today's marginal cost as MC0, and tomorrow's marginal cost MC1. The sum MC0+MC1 is the total marginal cost. In equilibrium MB equals total marginal cost. The optimal amount of risky sex, x^*, is plotted as the red vertical line. In panel (a), the benchmark parameters are $\kappa = 1.5$, price $p = 1.5$, $\lambda(x) = 1 - exp(-x)$, $\lambda_x = exp(-x)$, $\beta = 0.5$, $s = 2.0$, $y_0(s) = s$, $y_1(s) = 2s$. For our benchmark case (i.e., $s = 2.0$) we set $\gamma^+ = 0.2$ and $\gamma^- = 0.9$, and these survival probabilities increase with s multiplying them by a factor of 1.1 if $s = 2.2$ and by a factor of 0.7 if $s = 1.8$. In panel (b), we increase education to $s = 2.2$., and in panel (c) we decrease it to $s = 1.8$. See our discussion in Section 6.1.

Figure 8: The Risky Sex-Education Gradient: Evolution Across Stages of the HIV Epidemic



(a) Women



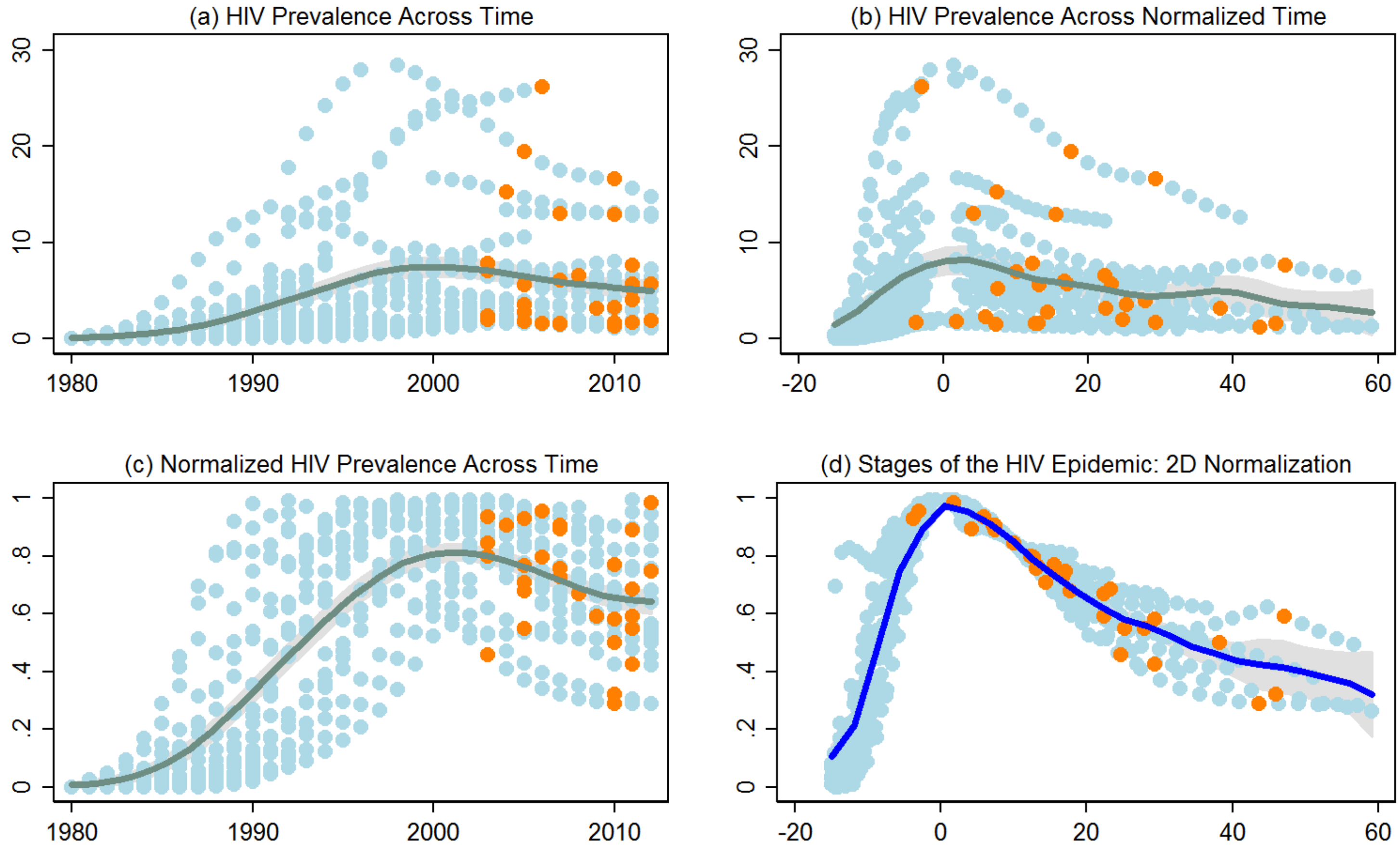
(b) Men

Notes: The HIV-Education gradient is plotted on the left vertical axis. The Partners-Education gradient and the Condoms-Eduction gradient are plotted on the right vertical axis. For each stage j we plot $\left(\gamma_0 + \sum_{j>0} \gamma_j \mathbf{1}_j\right)$. The specification we plot is with year controls. That is, for the HIV-Education gradient we use column 3 (8) in Table 5 for respectively women (top panel) and men (bottom panel). For the Partners-Education and Condoms-Education gradient we use column 3 (8) in Table 6 for respectively women (top panel) and men (bottom panel). Significance at 10%, 5%, and 1% is represented by, respectively, markers with open circles, markers with medium transparency fill, and markers with solid fill. We use a cubic spline for interpolation across stages.

Online Appendix

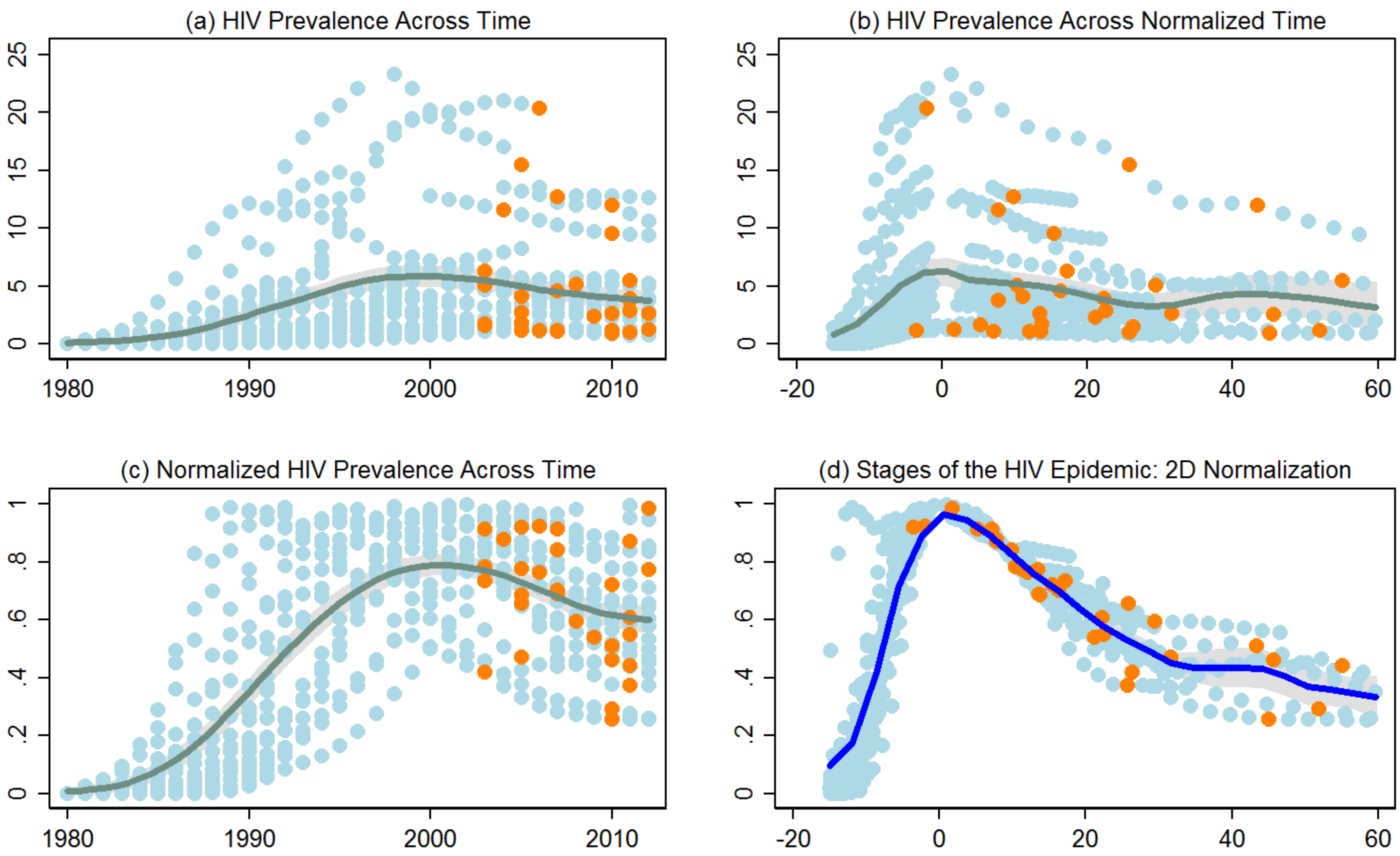
*Not for publication

Figure A-1: Stages of the HIV Epidemic for Women



Source: Location of SSA Countries (DHS Sample) on the 2D-Normalized Space at the Time of DHS Data Collection. Outcome of our 2D-normalization algorithm (Subsection 4.2) implemented using WPP, 2015, data. *Notes*: Each point in the scatterplot represents a DHS dataset.

Figure A-2: Stages of the HIV Epidemic for Men



Source: Location of SSA Countries (DHS Sample) on the 2D-Normalized Space at the Time of DHS Data Collection. Outcome of our 2D-normalization algorithm (Subsection 4.2) implemented using WPP, 2015, data. *Notes*: Each point in the scatterplot represents a DHS dataset.

Table A-1: The HIV-Education Gradient, Sexually Active Sample

HIV Status	(1)	(2)	(3)	(4)	(5)
Education	0.0048***	0.0123***	0.0107***	0.0042***	0.0039***
	(0.0009)	(0.0008)	(0.0013)	(0.0003)	(0.0004)
Education * Stage1		-0.0064***	-0.0047***	-0.0007	-0.0004
		(0.0010)	(0.0013)	(0.0004)	(0.0005)
Education * Stage2		-0.0129***	-0.0115***	-0.0029***	-0.0027***
		(0.0009)	(0.0014)	(0.0003)	(0.0004)
Education * Stage3		-0.0101***	-0.0081***	-0.0018	-0.0016
		(0.0020)	(0.0015)	(0.0015)	(0.0014)
Education * Stage4		-0.0067***	-0.0054***	-0.0019***	-0.0014***
		(0.0013)	(0.0016)	(0.0004)	(0.0004)
Male	-0.0267***	-0.0273***	-0.0272***	-0.0268***	-0.0269***
	(0.0039)	(0.0038)	(0.0025)	(0.0022)	(0.0021)
Age	0.0021***	0.0021***	0.0021***	0.0022***	0.0022***
	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0002)
Urban Area	0.0260***	0.0243***	0.0276***	0.0337***	0.0344***
	(0.0059)	(0.0059)	(0.0053)	(0.0038)	(0.0038)
Stage 1	-0.0018	0.0137***	0.0097	-0.0082***	0.0099***
	(0.0062)	(0.0047)	(0.0060)	(0.0012)	(0.0023)
Stage 2	0.0132	0.0573***	0.0676***	0.0010	0.0289***
	(0.0105)	(0.0122)	(0.0136)	(0.0029)	(0.0045)
Stage 3	-0.0114	0.0198	0.0330**	-0.0129**	0.0138**
	(0.0164)	(0.0149)	(0.0128)	(0.0061)	(0.0060)
Stage 4	-0.0027	0.0143***	0.0436***	-0.0183***	0.0056
	(0.0052)	(0.0040)	(0.0097)	(0.0025)	(0.0036)
Agricultural Share	-0.0034***	-0.0034***	-0.0035***	0.0027***	-0.0006
	(0.0003)	(0.0003)	(0.0004)	(0.0005)	(0.0004)
Output per Capita	-0.0000***	-0.0000***	-0.0000***	0.0001**	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.1128***	0.0918***	0.0554***	-0.1603***	-0.1992***
	(0.0101)	(0.0075)	(0.0098)	(0.0493)	(0.0282)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	329,205	329,205	329,205	329,205	329,205

Notes: All specifications use the "Sexually Active" subsample. he underlying econometric models are as specified in the columns of Table 3. Column (1) reports the tests results for the stationary specification. Columns (2) to (5) report the tests results for the non-stationary specification. We include the same set of controls and fixed effects as in our benchmark specifications in Table 3. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A-2: The Knowledge-Education Gradient: Women and Men Separately

(A) One Sexual Partner without Other Partners

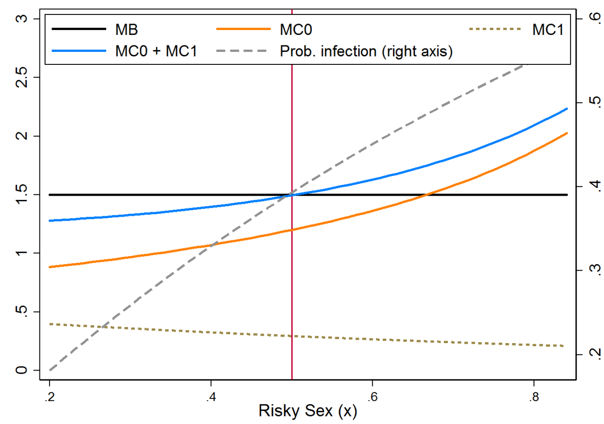
	Women					Men				
HIV Knowledge	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0135	0.0033	0.0005	0.0074***	0.0076***	0.0076**	0.0010	0.0013	0.0078***	0.0087***
	(0.111)	(0.110)	(0.866)	(0.000)	(0.000)	(0.001)	(0.326)	(0.580)	(0.000)	(0.000)
Education * Stage1		0.0126	0.0120**	0.0047	0.0046		0.0078***	0.0059**	0.0005	-0.0001
		(0.156)	(0.043)	(0.309)	(0.292)		(0.000)	(0.039)	(0.840)	(0.952)
Education * Stage2		0.0072**	0.0093**	0.0038**	0.0033		0.0091***	0.0086***	0.0028*	0.0017
		(0.046)	(0.049)	(0.045)	(0.210)		(0.000)	(0.001)	(0.088)	(0.285)
Education * Stage3		0.0094	0.0120	0.0018	0.0020		0.0057	0.0053	-0.0034	-0.0042
		(0.612)	(0.412)	(0.912)	(0.911)		(0.386)	(0.337)	(0.444)	(0.304)
Education * Stage4		0.0104	0.0134	0.0071	0.0062		0.0069**	0.0075*	0.0019	0.0007
		(0.732)	(0.648)	(0.812)	(0.830)		(0.044)	(0.066)	(0.424)	(0.759)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	213,907	213,907	213,907	213,907	213,907	167,894	167,894	167,894	167,894	167,894

(B) Always Use Condom During Sex

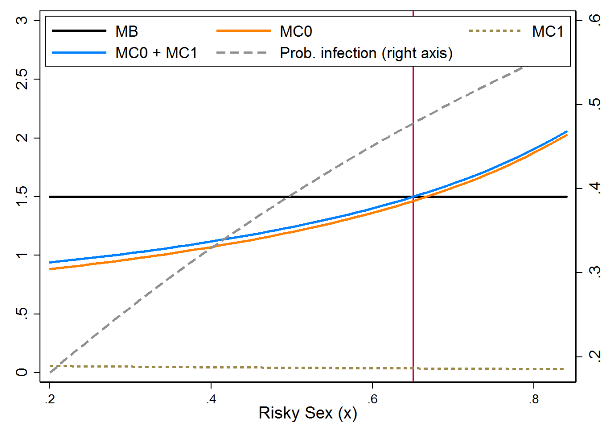
	Women					Men				
HIV Knowledge	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Education	0.0221	0.0194***	0.0169***	0.0196***	0.0196***	0.0126***	0.0132***	0.0141***	0.0187***	0.0181***
	(0.133)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Education * Stage1		0.0044	0.0073	0.0034	0.0039		-0.0006	-0.0007	-0.0036	-0.0029
		(0.801)	(0.593)	(0.809)	(0.754)		(0.935)	(0.929)	(0.638)	(0.668)
Education * Stage2		-0.0026	0.0003	-0.0017	-0.0025		0.0015	0.0017	-0.0036	-0.0035
		(0.616)	(0.953)	(0.648)	(0.572)		(0.639)	(0.637)	(0.236)	(0.246)
Education * Stage3		0.0009	0.0039	0.0008	0.0014		-0.0028	-0.0024	-0.0073	-0.0064
		(0.956)	(0.804)	(0.952)	(0.920)		(0.770)	(0.789)	(0.415)	(0.496)
Education * Stage4		0.0067	0.0089	0.0047	0.0033		0.0006	-0.0004	-0.0046*	-0.0047**
		(0.946)	(0.929)	(0.963)	(0.974)		(0.828)	(0.910)	(0.065)	(0.029)
Year-Country Dum.	No-No	No-No	Yes-No	No-Yes	Yes-Yes	No-No	No-No	Yes-No	No-Yes	Yes-Yes
Sample Size	213,763	213,763	213,763	213,763	213,763	167,800	167,800	167,800	167,800	167,800

Notes: In panel (A) we report the coefficients of a linear model where the endogenous variable is binary for *"Can you (the respondent) reduce the chances of getting HIV by having one sex partner who has no other partners?"*. In panel (B) we report the coefficients of a linear model where the endogenous variable is binary for *"Can you (the respondent) reduce the chances of getting HIV by always wearing a condom?"*. In both panels we include the same set of controls and fixed effects as in our benchmark specifications in Table 5. Standard errors are clustered at the country level using the wild cluster bootstrap from Cameron et al. (2008), and reported in parenthesis.* significant at 10%; ** significant at 5%; *** significant at 1%.

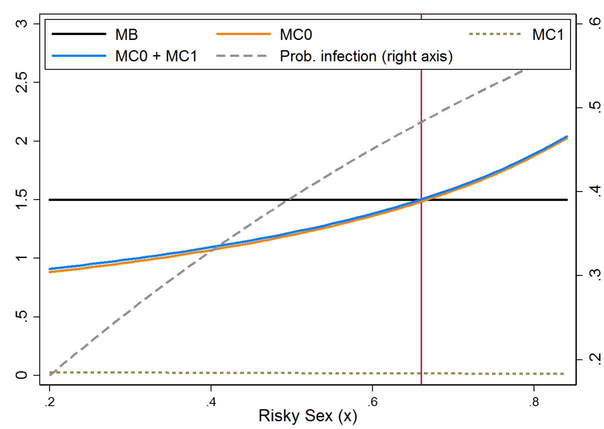
Figure A-3: Theoretical Interpretation: ART



(a) Benchmark



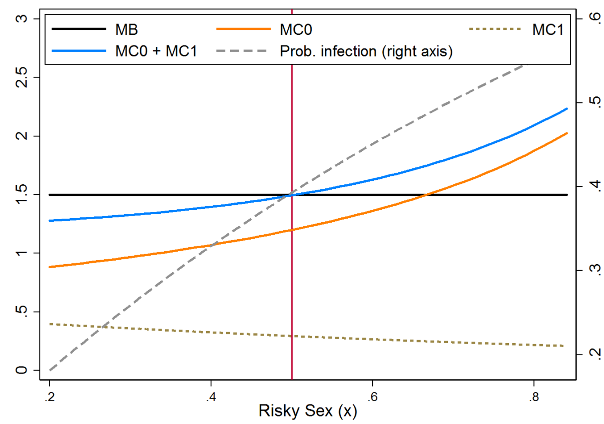
(b) Increasing Survival Probability (ART)



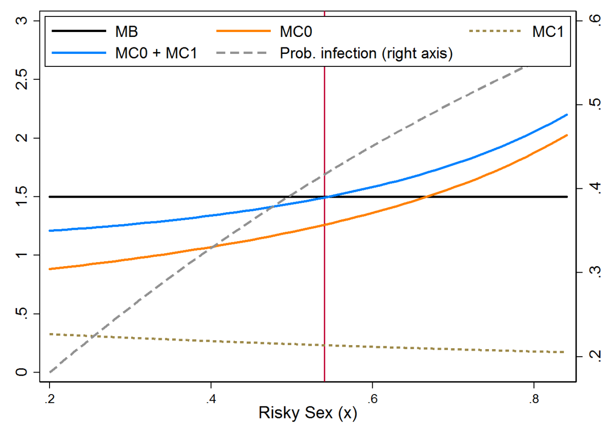
(c) Reducing Infectiousness (ART)

Notes: The horizontal axis denotes risky sex, x. All variables are plotted on the left vertical axis, except for the probability of infection $\lambda(x)$ that is plotted on the right axis. The marginal benefit (MB) is κ. The marginal cost through current consumption, $MC0$, is $u_c p = p/c$, and through future consumption, $MC1$, is $\lambda_x(\gamma^+ - \gamma^-)u(c_1)$. The sum of $MC0$ and $MC1$ is the total marginal cost. In equilibrium MB=MC. In panel (a), the benchmark parameters are $\kappa = 1.5$, price $p = 1.5$, $\gamma^+ = 0.2$, $\gamma^- = 1.0$, $\lambda(x) = 1 - exp(-x)$, $\lambda_x = exp(-x)$, $\beta = 0.5$, $s = 2.0$, $y_0(s) = s$, $y_1(s) = 2s$. In panel (b), we increase survival rates through ART, we set γ^+ 0.8. In panel (c), we decrease the HIV infectiousness, we set $\lambda_x^{new} = 0.5\lambda_x^{old}$.

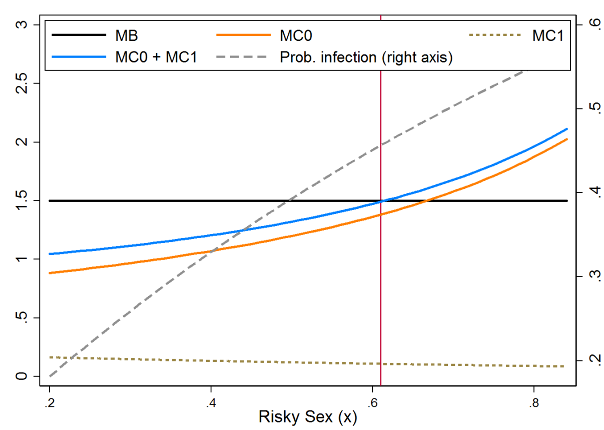
Figure A-4: Theoretical Interpretation: ART with Private Monetary Cost



(a) Benchmark



(b) Increasing Survival Probability (ART) with Monetary Cost



(c) Reducing Infectiousness (ART) with Monetary Cost

Notes: The horizontal axis denotes risky sex, x. All variables are plotted on the left vertical axis, except for the probability of infection $\lambda(x)$ that is plotted on the right axis. The marginal benefit (MB) is κ. The marginal cost through current consumption, $MC0$, is $u_c p = p/c$, and through future consumption, $MC1$, is $\lambda_x(\gamma^+ - \gamma^-)u(c_1)$. The sum of $MC0$ and $MC1$ is the total marginal cost. In equilibrium MB=MC. In panel (a), the benchmark parameters are $\kappa = 1.5$, price $p = 1.5$, $\gamma^+ = 0.2$, $\gamma^- = 1.0$, $\lambda(x) = 1 - exp(-x)$, $\lambda_x = exp(-x)$, $\beta = 0.5$, $s = 2.0$, $y_0(s) = s$, $y_1(s) = 2s$. In panel (b), we increase survival rates through ART, we set γ^+ 0.8, and also make agents pay for ART with a cost of 3.00. In panel (c), we decrease the HIV infectiousness, we set $\lambda_x^{new} = 0.5\lambda_x^{old}$ and also make agents pay for ART with a cost of 3.00.